Bank costs, structure, performance and regulation : evidence from the UK, 1981-1995.

Evripidou, Loukia Ch

Bank Costs, Structure, Performance and Regulation: Evidence from the UK, 1981-1995

By

Loukia Ch. Evripidou

Doctor of Philosophy in Finance

University of Durham

Department of Economics and Finance

2001

Abstract

Bank Costs, Structure, Performance and Regulation:
Evidence from the UK, 1981-1995

By Loukia Ch. Evripidou

Given the changes in the structure of the banking industry in the last two decades, the current study attempts to assess their impact on the costs and profits of UK banking industry, by use of a new data set covering the period 1981-1995. By utilising different methodologies we provide an insight into the robustness of the findings regarding different methodological specifications.

First, we identify economic and market trends, as well as regulatory alterations that might have stimulated the changes in the UK banking industry over the 1980s and the 1990s in the context of modern banking theory. Second, we assess scale economies and cost complementarities by estimating a cost function for UK banks. The findings indicate the existence of decreasing returns to scale and the lack of cost complementarities at the beginning of the data period. Also, following deregulation during the 1990s there is evidence of economies of scope in joint production. Third, by estimating an efficient frontier and by measuring the average differences between observed banks and banks at the frontier, we find evidence that X-inefficiencies account for 20 to 30% UK banking costs. Fourth, we test the relationship between market structure and profits. The evidence supports the structure conduct performance hypothesis and its 'quiet life' addendum. Finally, we investigate whether deregulation altered the costs, profits, and the efficiency of banking institutions. Evidence suggests that after deregulation there is increase in both bank costs and profits.

The material contained in this thesis has not been previously submitted for a degree in this or any other University.

Acknowledgements

I would like to thank my supervisors Mrs. Gioia Pescetto, and Professor Antonios Antoniou, for their valuable support, guideline and insightful comments through the course of my study.

I am grateful to my parents for their encouragement, as well as moral and financial support over the years. Without their valuable help the completion of this thesis would have been impossible.

I am also thankful to Dr. Sarantis Kalyvitis for his editorial comments. I would also like to thank my friends Sevasti Kaplanoglou and Charalambos Zevlaris for their constant encouragement and invaluable help. Finally many thanks to all the Faculty of the Department of Economics and Finance both at Durham and at Brunel University for their valuable assistance throughout the period of my studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The main motivation for the current study emerges from three factors. First, from the importance for the banking sector to the economy. Second, from the need to assess the effects of fundamental changes that occurred in the cost, profit, structure and performance of the banking industry in the UK over the 1980s and early 1990s. Third, from the lack of related empirical literature on the performance of the UK banking institutions. The results from this study can contribute to understanding the performance of banking institutions in the UK banking industry over a prolonged period of market and regulatory changes.

In most countries the business of banks did not change significantly from the late 1940s to the 1970s. This is certainly related to the strict regulatory framework followed by authorities in each country. In the 1980s, however, we have witnessed a revolution in both banking theory and banking practice. This was due to an international deregulation pattern that changed the nature of banking, and consequently its production and cost structure. Aiming at liberalising the provision of services and making easier the entry of new banks in the market, this deregulation pattern emerged mainly by two mutually reinforcing causes. The first one was increasing internalisation of the non-financial sector, and the second one was that the existing regulations were largely set up for the needs of the past, and hence no longer satisfied current financial needs.

These changes in the structure of the banking industry have put an increasingly sharp focus on whether they have really improved bank performance. Banks provide liquidity, payments, safekeeping and portfolio services for depositors, and intermediate their funds into investments and working capital resources. They also ensure a smooth functioning of the economy by allowing financial and real resources to flow relatively freely to their highest return uses. Hence, it is important to determine whether these changes have altered positively the efficiency of banking institutions. In particular, if banks become more efficient, then we expect improved profitability, greater amounts of funds intermediated and lower prices and ameliorated quality service for consumers. Also, if they are more efficient, then they become more viable, and the possibility of failure or being subject to take-over, decreases substantially.

Another important issue arising from deregulation concerns changes in the market structure of the industry. The effectiveness of banks in servicing deposits and credit needs is related to the structure of the banking industry. Two basic policy issues evaluate which banking structure could best serve the economy in terms of both cost and availability: (a) efficiency and (b) minimisation of failure. Efficiency is related to competition; in a more competitive environment, industries become more efficient. On the other hand, under monopolistic conditions, firms that are inefficient can still operate. In more concentrated markets, firms with higher market power may not gain from non-competitive pricing in the form of profits, but by benefiting from a more relaxed environment in which less effort is put on maximising cost efficiency [Berger (1995, 1997), Berger and Hannan (1998)]. Accordingly, due to lack of competition inefficient banks may still operate. Therefore, the objective of a failing proof banking industry is being served under monopolistic condition, whereas a competitive industry provides favourable conditions for more efficient firms. Using evidence from the effects of changes of the market structure on banks' performance, regulators are able to pursue optimal policies. If bank performance is not affected by alterations in the structure of the industry, then regulations are not necessary. In the opposite case, regulators must intervene to maintain the well functioning of the industry and the economy.

Until now, the bulk of empirical work has concentrated on the US banking industry, and little work has been done outside this region. However, US banking has a different structure from other banking industries. In Europe, banks tend to be multi-branch and universal, whereas branching (in some states) and universal banking in US are prohibited. Thus, any information obtained from US studies is of limited importance to other countries, and more research should be undertaken in the area of European bank cost and profit structure.

Given the above, the objectives of the current thesis are, first, to assess UK banks' ability to realise economies of scale and cost complementarties in joint production by modelling the structure of costs and production for a multi-product institution. Second, to appraise the X-efficiency of UK banks by estimating an efficient frontier and by measuring the average differences between observed banks and banks at the frontier. Third, to examine the market structure and profit relationship in attempting to evaluate which market structure characterises the UK banking industry. Fourth, to determine the changes in UK bank profits, cost and efficiency following certain regulatory changes.

In Chapter 1, we try to identify economic and market trends, as well as regulatory alterations that stimulated the changes in the UK banking industry over the 1980s and 1990s. With the beginning of 1980s, British banking started to change. Improvements in technology introduced new facilities to bank services, and constituted a major part of the improvement process in banks' efficiency. Balance sheet structure changed rapidly with the introduction of off-balance sheet activities, product innovation, and banks' movement into security trading. Furthermore, two new Banking Acts (1979, 1987) provided the industry with a detailed framework of prudential control. Finally, the major trends in banking regulation since 1987 are defined by the international convergence on common minimum capital standards of capital adequacy and the trend of regulatory policy towards a common European standard.

Chapter 2 is of a theoretical nature and analyses some of the theory of modern banking firm in the context of the current thesis. Recent studies on banks emphasise their role in transforming illiquid assets into liquid liabilities, [Diamond and Dybving (1983), Freeman (1988)], minimising transaction costs by acting as delegated monitors of investors, [Diamond (1984,1989,1996), Chant (1987)], and providing payments and portfolio services. The interruption of these services may be costly to the economy as the production process may be severely affected and assets liquidated prematurely. In such a case, markets fail to provide the equilibrium quantity of services at the market price. Accordingly, banks are regulated in order to ensure the adequate provision of these services and to minimise the possibility of market failure. The next part of this chapter concentrates on the definitional and measurement problems of banks' input and output. Given the nature of banks' activities, the literature has exposed considerable difficulties defining what constitutes bank input and output, and has employed different approaches. We provide an overview of these approaches, together with a summary table of the input and output measures used in the bank cost literature. Finally, key competitive issues are examined, with an extensive discussion of scale and scope economies, X-efficiencies, and market structure relationship in banking.

The objective of Chapter 3 is to assess UK banks' ability to realise economies of scale and cost complementarties in joint production by modelling the structure of costs and production for a multi-product institution. We employ two approaches in defining input and output, and we use a new data set to check the robustness of the results to different input and output specifications. In particular, scale economies suggest that larger firms produce at lower average cost than smaller firms, whereas cost complementarities in the production of two products exist if the marginal cost of producing one output decreases when the production of the other increases. This evidence is important for decision-making by banks and by regulatory authorities. After reviewing the alternative methodologies employed by the literature, we opt for the translog cost function; this cost function is preferred in studies of bank performance because it allows, first, the estimation of a U-shaped average cost curve, and, second, the derivation of scale and scope economies [see e.g. Mester (1993), Jagtiani *et al.* (1995), and Rogers (1998a,b)]. We evaluate economies of scale as the sum of the first-order partial derivatives of the estimated cost function, with respect to output [see, among others, Fields *et al.* (1993), Jagtiani *et al.* (1995) and Jagtiani and Khanthavit (1996)], and cost complementarities as the second order derivatives of the cost function with respect to output [see, among others, Mester (1987), Allen and Rai (1993), Jagtiani *et al.* (1995), and Rogers (1998a)].

In Chapter 4 we test the X-efficiency of UK banks by estimating an efficient frontier and by measuring the average differences between observed banks and banks at the frontier. Following Leibenstein (1966), X-efficiency is defined as the effect of differences in managerial ability to maximise revenue or minimise cost, and accounts for differences in costs that cannot be explained by differences in scale or other observable characteristics. Recent literature suggests that X-efficiencies account for 20% or more of cost in banking whereas scale and product diseconomies only account for less than 5% of cost [Berger *et al.* (1993)]. In this chapter we first provide a review of the different approaches for measuring X-efficiency and their empirical results. Next, based on its advantages over other frontier methods we employ the stochastic *econometric frontier approach* [Berger and Mester (1997a), Berger and DeYoung (1997), Bauer *et al.* (1998)]. To test the effect of this assumption on the robustness of the results we estimate three alternative models with different distributional assumptions. The results suggest that the efficiency estimates are fairly robust to differences in methodology.

In Chapter 5 we examine the market structure and profit relationship in the UK banking industry by including direct measures of both market structure and efficiency. Many studies in the banking literature (especially in the US) have examined the connection between profitability and measures of market structure (concentration or market share), and have found a positive statistical relationship. There are two interpretations for this finding: the first is the *market power hypothesis*, and the second is the *efficient structure hypothesis*. The *market-power hypothesis* suggests that firms in concentrated markets exercise market power in pricing and earn supernormal profits [Shepherd (1982)]. The *efficient structure* paradigm links concentration to high profitability through efficiency [Demsetz (1973)]. Following Berger (1995, 1997) we establish structural forms for the efficient structure and market power hypotheses, and we derive a reduced-form model that nests all related hypotheses, which helps to determine what form of market structure characterises the UK banking industry. An innovation of the current chapter is the utilisation of different efficiency measures to test the sensitivity of the results. We further test the existence of the 'quiet life' hypothesis. As suggested by Hicks (1935), the reduction in competitive pressure in concentrated markets may result in higher cost per unit of output due to slack management.

In Chapter 6, our goal is to determine the changes in UK bank profits, cost and efficiency following certain regulatory changes for the period 1983 to 1995 by examining the regulatory changes in the UK banking system. The motivation for this test comes from the global deregulation of the banking industry that has taken place during the 1980s and has had an important effect on the performance of banking institutions. However, the quantification of these effects has only been examined by a few studies, and no evidence is available for the UK banking market. In the first part of the chapter we review the relevant literature, and we describe the major regulatory tools in the banking industry. Next, we concentrate on the studies that quantify regulatory changes on banks' performance. Following Humphrey and Pulley (1997) and Berger and Mester (1997, 1999) we divide the data into three time periods, namely pre-regulation, concurrent, and post deregulation. For each period we estimate a profit and a cost function and we separate internal bank-initiated adjustments from external changes in the banking environment. This separation enables us to identify the major impact of changes in regulation on bank profits and costs.

CHAPTER 1: Recent developments and structural changes in the UK banking system

1.1 Introduction

The aim of the current chapter is to identify economic and market trends, as well as regulatory alterations that stimulated the changes in the UK banking industry. The face of the British banking has changed substantially during the past three decades. Prior to the 1970s it was dominated by a cartelised oligopoly. By the beginning of the 1970s legal, political and banking environment became much more favourable to the encouragement of free competition in banking.[1] Banks faced new competitors and new forms of competition were developed. This intensifying competition increased the pressure for innovation, whereas improvements in technology lowered the corresponding costs of innovating behaviour by banks. Furthermore, the macroeconomic climate experienced by the banking industry throughout the 1970s and 1980s was much more volatile than in the 1950s and 1960s. The rise of interest rates, and inflation, accompanied by increased budget deficits prompted banks to reassess their attitudes towards risk and uncertainty, and augmented the economic impact of traditional regulations on banks [Gardener and Molyneux (1994)].

Banking industry in the UK, prior to the Banking Act 1979, was practically unregulated. In practice, governmental control was ensured with the combination of selective legal restrictions on certain banking activities, together with the Bank of England's control over the structure and operation of the banking industry. The banking crisis of 1973-74 highlighted the need of a detailed framework of prudential controls for the banking market. As a result, two Banking Acts were brought into being, Banking Act 1979 and 1987. The main trends in banking regulation since 1987 are defined by international convergence on common minimum standards of capital adequacy and the increasing Europeanistation of regulatory policy. The purely domestic changes in the regulatory structure have been few and relatively modest.

[1] Bank of England Quarterly Bulletin, September 1983.

These changes in the structure of the UK banking industry give rise to the question of their effects on banks' performance. This gives us the motivation for the current study. As this study concentrates on estimating the performance of UK banks for the period 1980 to 1995, this chapter provides an insight to all of the changes in the industry following the introduction of the Banking Act of 1979 until the Bank of England Act of 1998. The first part deals with the market developments from early 1970s till mid 1990s. The second part of this chapter deals with the regulatory developments both in the UK and EC, for the same period.

1.2 Market developments

Four decades ago, the British banking system was dominated by a cartelised oligopoly consisting of the London clearing banks and their associates in Scotland and Northern Ireland. By the beginning of 1970s, the legal, political and banking opinion became much more favourable to the encouragement of free competition in banking and much more hostile both to cartels and to similar counter-competitive techniques of monetary policy[2]. Drastic deregulation followed, where in response to official prompting, clearing banks changed the form of their published accounts. At the same time, again with official approval, extensive mergers substantially reduced the number of clearing banks.

During the ensuing competitive decade (and though partly interrupted by some re-imposition of direct controls during 1974-80) the banking industry started to change with the introduction of new facilities for its customers. For personal customers, credit cards and cash dispenser became readily available, so that by 1982 nearly 80% of bank current account holders held at least one such card. Automated credit transfer and direct debit facilities became available, while personal customers also gained easier access to a more complete array of instalment facilities. Corporate customers had access on a growing scale to the wholesale market, whether as lenders or borrowers of short-term money. The cost of such customers was thereby reduced [Bank of England Quarterly Bulletin (September 1983)].

[2]For example restrictive lending ceiling applied to banks and depository institutions

In the beginning of the 1980s, a radical change in the structure of the UK and of the international financial sector took place. This was due to two factors. First, following the international trend large financial institutions provided a wide range of financial services, and consequently, specialisation in particular areas or activities was discouraged. After the considerable changes in supervision during 1985, the largest UK banks concentrated on providing both commercial and investment banking globally, and move towards universal banking. On the other hand medium and small banks continued to focus on more traditional activities and less ambitious investment banking strategies [Banking Act Report (1985/1986)]. Second, pressure for change grew considerably due to the increase in competition within the banking industry. Banks faced new competitors and new forms of competition were developed.

1.2.1 Market Overview from late 1970s to mid 1990s

During late 1970s and early 1980s entered a period of recession.[3] At that period UK banks faced a continuous decline in their capital ratios. In 1982 there was a sign of rise, in line with the recover of economic activity. However, following the Finance Act 1984, in March 1984, there was a decline in post-tax profits caused by substantially higher effective tax rates. Most banks increased provisions through transfer from existing reserves, which resulted in deterioration in their capital ratios. The banks affected have taken measures to recover their capital, mainly by issuing subordinated loan stock and to a less extend from issuing of capital [Banking Act Report (1984/1985)].

During the following years there was an increasing pattern in the banks' capital ratios and profits, and this was mainly emerge due to changes in their structure and strategies, and due to the decline of corporate tax rate. Changes in banks' structure and strategies stem mainly from the increase in the number of off-balance sheet activities, product innovation, increasing participation in security markets and banks' movements into investment banking and securities trading and distribution, both in the UK and overseas. Diversification and increasing involvement in the domestic and international security markets raised a number of questions concerning the assessment of risk and the quality of earnings. On the one hand, diversification can be thought as a tool in reducing bank's overall risk by achieving a greater balance in portfolio exposure. On the other hand, trading losses incurred in investment banking and securities operations can significantly damage the earnings even of largest institutions [Banking Act Report (1986/1987)]. A clear example for this concern came forth some years later. In 1995 the merchant bank Barings collapse after experiencing large losses on futures and option trading by a Singapore subsidiary. Because of the risk faced by banks due the utilisation of the aforementioned instruments, the Bank of England as well as the EC supervisory body, issued several regulatory notices for the matter in concern (a detailed analysis is given in the next section).

Another point of notice for this period involves the considerable concerns raised for those banks with concentration of lending to major debtor countries. During 1982 the number of countries experiencing debt-serving difficulties, increased. Although the immediate debt problem of these countries tended to diminish trough time due to improvements in their external positions, the resolution still depended largely on their domestic economies and world growth. As a result in 1987 large British banks increased their bad debts provisions to around 25%-35% of their exposures to problem countries. Nevertheless, this increased in the level of provision, led to a sharp fall in profits [Banking Act Report (1987/1988)].

Apart from increases in debt provisions, banks that diversified into securities markets during the past couple of years faced a further problem. Banking business closely related to stock market activities remained severely affected by market's difficulties following the sharp fall of prices in October 1987. The majority of banks' securities markets operations, both equities and fixed interest, were significantly affected by strong competition, by lower levels of turnover in the market and by narrower margins. Nevertheless, the experience was not uniform and securities business continued to be profitable for certain banks, particularly large ones. Low level of activity and reduced value of funds under management also affected investment management activities [Banking Act Report (1987/1988)]. The adverse effects of 1987 were balanced, to some extent, from continuing rapid growth in lending to personal and to corporate sector. Figure 1-1 presents the changes in pre- and post-tax

[3] International economy entered a period of recession after the oil shock in 1979.

profits for large UK banks[4] over the period 1982 to 1995, and Table 1-1 summarises the large UK banks capital ratios.

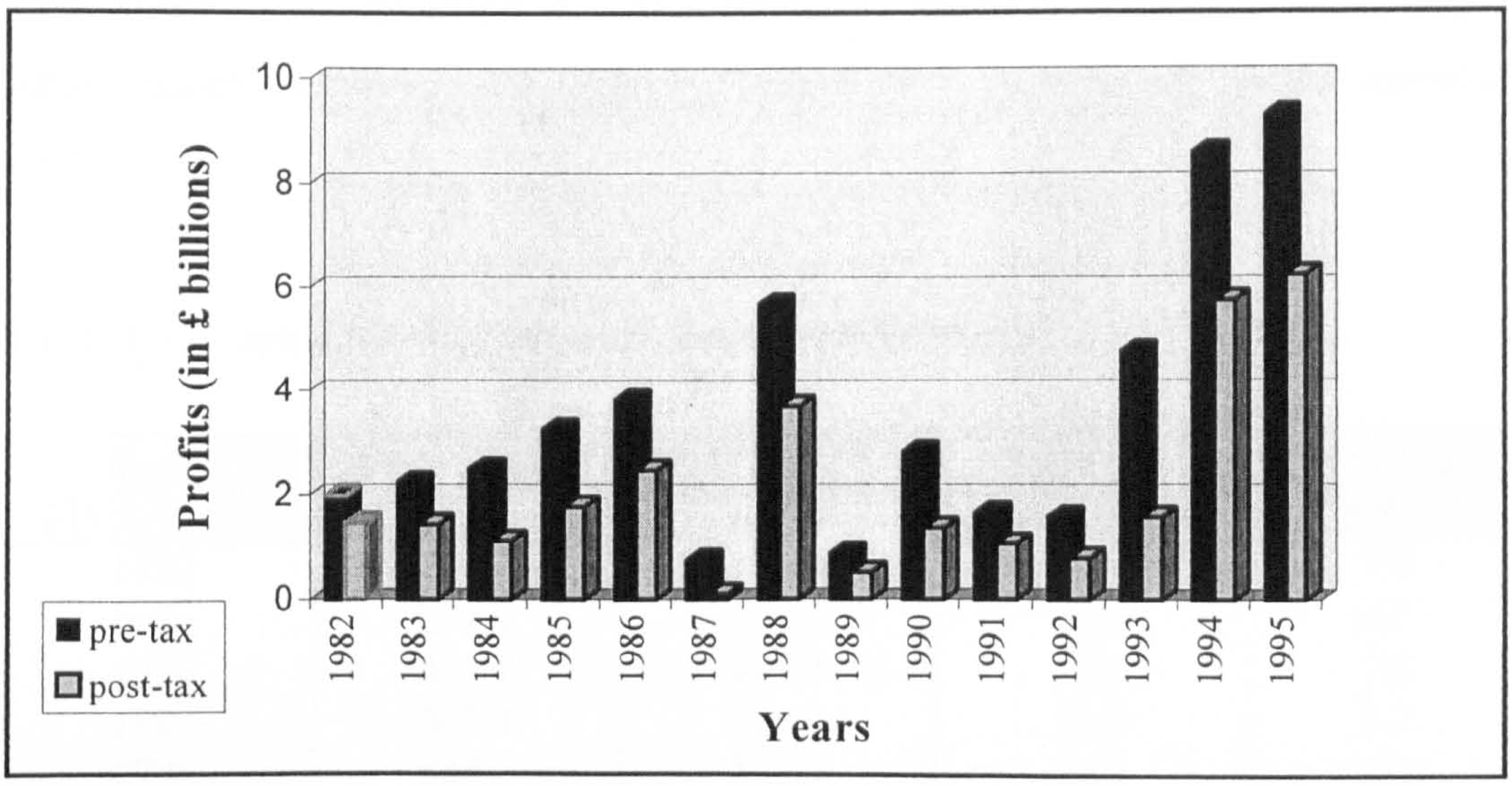

Figure 1-1: Pre- and post-tax profits (£ billions) of the UK largest banks for the period 1982-95 (Source: Bank of England, Banking Act Reports 1982-1996).[5]

As it can be seen from the Figure 1-1, there is a substantial upward movement in the profitability of large UK banks in 1988, the year after the allowance for large provisions against the problem of country debt. The recovery of their profit rates diminished their reliance on capital issues as a means of increasing capital and reserves. Furthermore, the strength of their retained earnings and the raising of new equity improved banks' overall capital ratios. However, tighter monetary policy in 1989 led to deterioration in the property markets and the corporate sector of the economy, while several companies started facing financial difficulties. Due to worsening economic conditions, signs of difficulties appeared in small and medium-sized companies with corporate insolvencies reaching near record levels in 1989 [Banking Act Report (1989/1990)]. To maximise profits large banks place emphasis on cost control and on less capital-intensive income sources. Moreover, the decision

[4] UK largest banks as defined by the Bank of England are Barclays, Lloyds, Midland, National Westminster, Standard Chartered, Bank of Scotland, The Royal Bank of Scotland and TSB.

of large banks to add substantial new provisions against problem country debt, as well as domestic commercial bad debt charges (sometimes up to 80%), in conjunction with general economic conditions, resulted in the fall in profits. In an attempt to lessen the effect of economic deterioration, banks improved the monitoring of doubtful debts, within branch networks, and upgraded information systems and credit appraisal techniques.

Table 1-1: Capital ratios of large UK banks in £ billions.

Year	Total assets	Weighted assets	Adjusted capital base	Risk asset ratio
1982	234.4	179.8	13.7	7.6
1983	271.1	208.0	16.9	8.1
1984	315.7	245.4	18.6	7.6
1985	303.1	236.0	23.0	9.7
1986	334.6	351.7	26.3	10.5
1987	345.2	267.2	26.3	9.8
1988	392.6	316.1	32.0	10.1
1989	461.7	377.1	34.7	9.2
1990	471.1	374.9	34.5	9.2
1991	476.8	364.8	35.5	9.7
1992	535.6	378.5	37.4	9.9
1993	593.0	381.2	41.3	10.8
1994	711.4	424.4	48.5	11.4
1995	809.7	471.1	51.1	10.9

Source: Bank of England Reports 1982 -1996

At the beginning of 1990s the UK economy swung into recession and uncertainty was triggered by the events in the Middle East. The rate of growth of banks' domestic lending dropped, whereas the bad debt provisions made by banks against their UK portfolios increased. One of the major reasons for these high provisions was that due to financial liberalisation of the last decade, many borrowers became more geared and, thus, more vulnerable to combination of high interest rates and weakening demand. Banks' earnings were also affected by changes in interest rates. In the past, banks benefited during periods of higher interest rates from the existence of substantial non-interest-bearing elements in their deposits. However, the increasing competition in personal savings, both between the banks and between them and other financial institutions (e.g. building societies and life assurance companies) generated

[5] Data covering the period prior to 1982 is not available.

the need for new products offered to customers, such as a variety of interest-bearing accounts. During the 1970s the non-interest-bearing sterling deposits had reached 50% of total deposits for many of the large banks, whereas in the 1990 they had fallen to 15% as more and more depositors switched to interest-bearing accounts. In an attempt to alter the interest rate exposures arising from on-balance-sheet items, banks developed a range of off-balance-sheet instruments (e.g. the use of swaps to match floating-rate funding and fixed-rate lending) [Banking Act Report (1990/1991)].

Throughout the recession, operating profits of large banks were boosted by cost cutting and efficiency measures. These measures were taken in an attempt to influence operating profitability in the recession period, and as a response to competitive pressures, which had intensified after 1980s deregulation. The result of this re-structuring was a sharp reduction in the number of employees. The only factor that increased costs was the continuing heavy investment in information technology undertaken by most banks in conjunction with higher depreciation charges. The latter came forth in a desire for long-term benefits and by increased automation of labour intensive activities (credit assessments, data processing and storage of customer information). Computerisation also helped banks improve services by expanding their scope to match products more closely with demand. Furthermore, large banks sought other sources of operating income less vulnerable to the recession: for instance, non-interest income, being a form of income that requires less capital, grew substantially as a proportion to total income and contributed to the increase in operating profits. Major sources of growth were foreign exchange and money markets trading commissions and the penetration in life assurance [Banking Act Report (1991/1992)].

Recovery came forth in 1993, and hereafter there is a continuously increasing pattern in banks' profitability and capital ratio, albeit during 1995 the steady asset growth of large banks led to a slight decrease in their capital ratio. However, as can be seen from Table 1-1 the capital ratios of large banks, remained comfortably above the 8% Basle minimum capital (more details on the Basle requirements are given in the next section) for the last ten years. During 1993 banks' operating profits grew due to non-interest income, where as interest income remained depressed due to stagnation in lending and the shift of some banks' balance sheets into lower risk and lower margin business. Moreover, more stable economic and financial conditions led to a fall in the overall bad debt charges [Banking Act Report (1992/1993)].

Operating profits further improved in 1994. Nevertheless, the improvement in profitability was almost entirely due to reduced bad debt provisions together with significant write-back of provisions against problem country debts at some banks. There was only a slight increase in the interest income, where as non-interest income was lower than in 1993, after increasing for the previous eight years [Banking Act Report (1993/1994)]. Profits continued to rise in 1995 as well. UK banks benefited from the recovery in loan demand during 1995, and this produced an increase in net interest income. Furthermore, non-interest income increased as well resulted mainly from dealing profits [Banking Act Report (1994/1995)].

1.2.2 Small banks

During the 1980s medium and small UK-banks record grew in line with the banking sector. Nonetheless, towards the end of the decade, increasing competition from larger banks placed substantial pressure on their business activities. Moreover, with the introduction of Banking Act 1987[6] small banks faced heavy cost in terms of satisfying minimum criteria for authorisation. During the recession, the deterioration in the UK small-banking sector became clear. These banks were traditionally involved with the property market, which left them highly exposed to the effects of the recession, as the ability of personal and small-business' sectors to service debt was reduced, and as the property market decline. The Bank of England kept these banks under review for the whole period under consideration, and helped them re-order their affairs or wind down business in an orderly manner [Banking Act Report (1992/1993)]. More stable economic and financial conditions in 1993 eased financial pressures on these institutions [Banking Act Report (1992/1993)].

[6] A detail analysis of the requirements of the Banking Act 1987 is given in section 1.3.

1.2.3 Banks' portfolios

The composition of banks' loan portfolios changed considerably during 1980s. Since widespread international debt problems emerged in 1982, new syndicated lending to sovereign and overseas borrowers declined. Many of highly rated corporate borrowers began looking for finance directly to bond markets rather than to banks. On their part, banks also placed greater emphasis on managing the shape of their loan portfolios, with some swapping from international debt and ceding other types of loans in an attempt to reduce credit risks. Furthermore, banks tried to increase the liquidity of their portfolios with securitisation.

The domestic personal sector has been an important area for the UK banks. Towards the second part of 1980s residential loans grew substantially due to strong demand for housing finance. However, banks' reliance on personal borrowing brought them into competition with building societies. Several banks announced significant structural change in their expense pattern in terms of interest bearing accounts [Banking Act Report (1989/1990)]. During the second part of the decade, corporate lending increased as well, mainly because of the increase in take-overs and merger activities. The increase in personal and corporate lending balanced by a decrease in the financial sector and overseas lending, leaving the total value of loans within the banking industry relatively unchanged for a period of five to six years. This can be seen from Figure 1-2, which presents the total loans and deposits of the banking industry for the period 1979 to 1995 are presented.

In the light of the difficult trading conditions in the beginning of 1990s, banks became more cautious in their assessment of the creditworthiness of borrowers. As already mentioned, most banks increased bad debts provisions on domestic lending due to increased company insolvencies in property, retail, and service sector. Moreover, major foreign banks restricted their UK lending activities and re-focused on areas of comparative advantage, such as trade finance and other country home business. In general, with the beginning of the new decade banks' lending to UK residents decreased sharply, reflecting also monetary policy tightening which had begun in 1989.

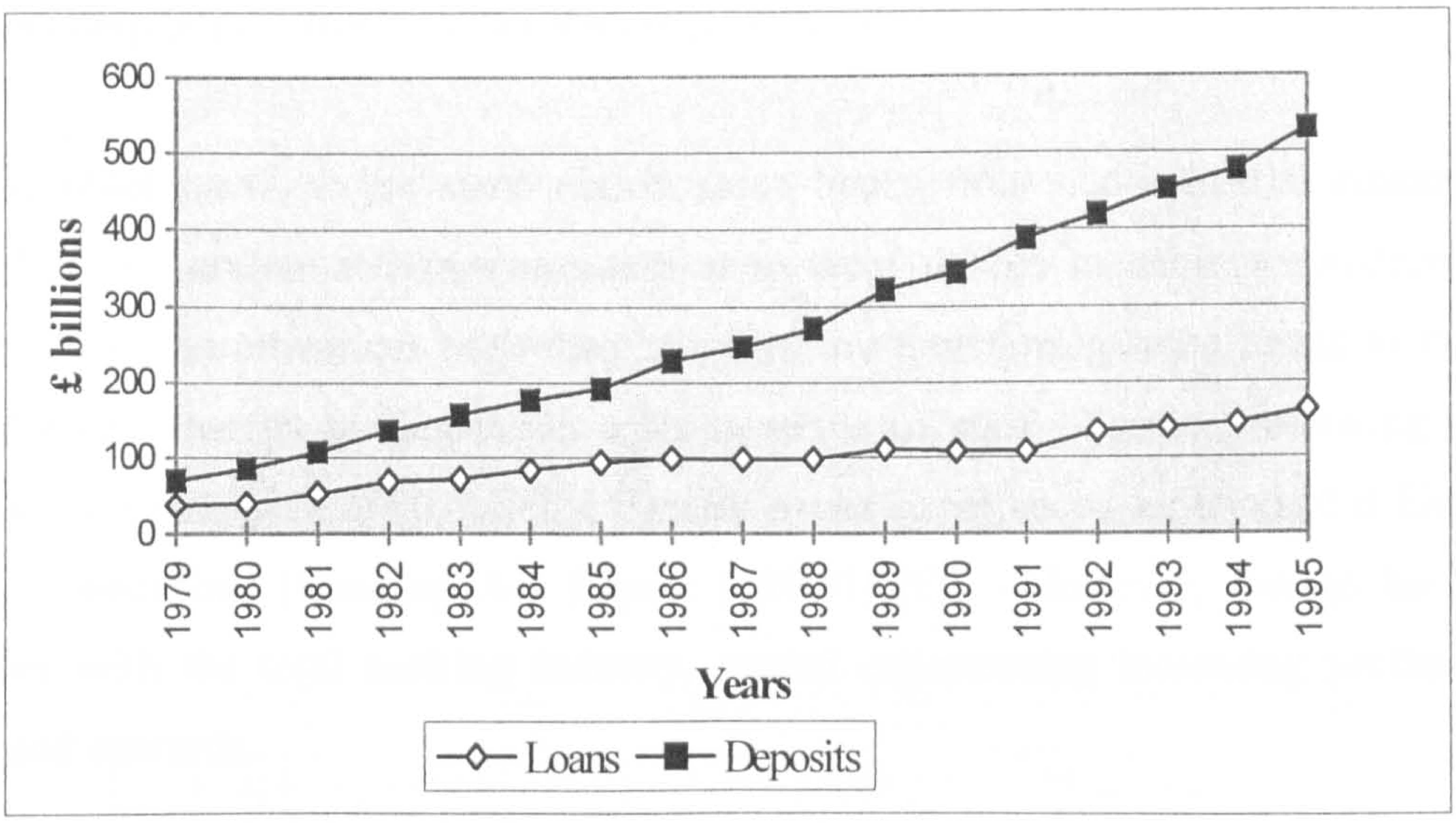

Figure 1-2: Total Deposits and Loans of the UK banking industry (£ billions) for the period 1979-1995 (Source: Datastream International).

As can be seen from Figure 1-2, although total deposits continued their increasing pattern, total loans remained approximately constant from 1986 to 1992. Following the improvements in the economic environment in 1993, total loans within the industry started growing, and continued following increasing pattern.

1.2.4 International developments

Apart from domestic changes, international developments also affected banking activities. Towards the beginning of 1990s, both European and non-European banks were reviewing their activities in the EU in accordance to the Single Market. In the light of the Second banking Directive (a detail analysis on the Directive is given in the next section), some non-European institutions were interested in opening subsidiaries in the UK, whereas several UK banks opened additional branches or established subsidiaries throughout Europe. Some overseas bank took more drastic steps in banking or related financial services through selective acquisitions. As a result several well-known British financial institutions came under the control of large foreign companies[7]. [Banking Act Report (1989/1990)]

On the other hand, at the same period some banks from non-industrial countries ceased their London activities, because they were unable to achieve satisfactory returns, whereas others cut back their activities by transferring some assets to their parent banks thereby cutting down costs in terms of staff. Further, following the Middle East events in 1990, Middle Eastern banks experienced a period of difficult trading conditions [Banking Act Report (1990/1991)]. However, foreign banks, together with the total banking industry, started experiencing increasing profits in 1993 and onwards.

Furthermore, banking industry in the 1990s was characterised by increased competition, by building societies, insurance, and investment companies. At the same time personal and corporate customers became more demanding, and were no longer willing to deposit money at a low rate of return. Hence, banks became interested in establishing better and closer relationships with their depositors, and in attracting more customers. As a result, a voluntary Code of Banking and Practice was issued, which came into force in March 1992. This set out minimum standards of good banking practice for dealing with personal customers. A revised edition was issued in February 1994 and included a commitment to provide advance notification to customers for changes to charges and interest rates.

[7] For example the national Australia Bank purchased the Yorkshire Banks, and the Bank of Yokohama purchased the majority of interest in the Guinness Mahon.

1.3 UK and EC Regulations

Prior to the Banking Act 1979, banking industry in the UK was formally unregulated. There was neither clear definition of a bank nor any statutory control over the use of banking names and description. In practice, governmental control was ensured by the combination of selective legal restrictions on certain banking activities and the Bank of England's control over structure and operation of the banking industry. The banking crisis of 1973-74 highlighted the need of a detailed framework of prudential controls for the banking market. As a result, two Banking Acts were brought into being, Banking Act 1979 and Banking Act 1987. Under these regulatory frameworks, unauthorised conduct of deposit-taking business has been prohibited. Moreover, the new legislation expanded and legitimised the regulatory discretion of the Bank of England. Another point of notice is the creation of new entry barriers in the industry, to prevent unnecessary competition and to reduce the contestability in the banking market to the benefit of established institutions. Nevertheless, the main trends in banking regulation since 1987 are defined by international convergence on common minimum standards of capital adequacy and the increasing Europeanistation of regulatory policy. The purely domestic changes in the regulatory structure have been few and relatively modest. In this section a review of the changes in the regulation is presented, starting with the beginning of 1970s.

In May 1971, the government in an attempt to combine an effective measure of control over credit conditions with greater scope for competition and innovation in the banking industry, issued a policy paper *'Competition and Credit Control'* (C.C.C.) that was implemented in 16 of September of the same year. The introduction of C.C.C. abolished the collective agreement of clearing banks on the interest rates and in 1972 the bank rate was replaced by the Minimum Lending Rate. Furthermore, government policy encourages overinvestment through a rapid expansion of property developments. However, due to the collapse in the property markets and the sharp deterioration in the general economic situation during 1973, it was clear after two years that reduced asset requirements led to large increase in the money supply. Therefore, the government in the 17th of December in an attempt to overcome the problem, presented its proposals for a new Development Land Tax on commercial property and the introduction of a 'corset'. In particular, the corset was a special deposit scheme that was intended to impose ceilings on the growth of banks' interest-bearing eligible liabilities with penalties in the form of non-interest bearing special deposits. However, two days after the announcement, in the 19th of December, a banking crisis broke, as a number of banks were unable to replenish their resources. This came to be known as the secondary banking crisis of 1973.

Following the secondary banking crisis and the EU First Banking Directive of 1977, the Banking Act was passed in 1979. The purpose of the EU Directive was to require from Member States to set up systems for authorising and supervising credit institutions. Accordingly, the Banking Act 1979 required the classification of deposit takers as either 'recognised banks' or 'licensed institutions'. The Act set up a set of minimum conditions that a bank should fulfil in order to be classified as a recognised bank:

- A 'high reputation and standing in the community'.
- Supply of 'either a wide range of banking services or a highly specialised banking service'.
- Business performance 'with integrity and prudence'.
- Direction under at least two individuals.
- Meets the minimum net asset requirements compromised in the Act but also to be considered appropriate by the Bank of England.

The licensed depository institution had to fulfil the latter three conditions, and instead of the first condition the following requirement should be satisfied:

- Demonstrate that all directors, controllers and mangers are 'fit and proper' to carry out business.[8]

The Act came into effect in October 1st 1979, and although it changed the process by which UK banking industry was regulated, the operation of the system still depended to a large degree on moral suasion.

Further to the Banking Act, the Bank of England (hereafter the Bank) in an attempt to clarify some important issues in banking process, in 1979, circulated three consultation papers within the banking community on important aspects of banks and

[8] Schedule 2 of the Banking Act 1979 amended in the Banking Act 1987.

other deposit-takers operations. The first paper, entitled '*The measurement of capital*', explains the basis upon which the Bank of England appraises capital adequacy of deposit taking institutions. There are two methods for assessing the adequacy of capital: namely the gearing ratio and the risk asset ratio. The gearing ratio is the ratio of current liabilities to capital resources, excluding that part of capital utilised to finance infrastructure and other non-banking activities. The risk asset ratio refers to the risk of losses generated by the use of institution's assets to capital available to finance such losses.[9] The objective of these two capital ratios is, first, to ensure that capital of an institution is regarded as acceptable by its depositors and other creditors (gearing ratio), and second, to test the adequacy of capital in relation to the losses which may be sustained (risk asset ratio). The definitive paper was issued in September 1980 and the Bank considered this paper as the basis on which it measures the adequacy of the capital of authorised institutions. (Composition of gearing and asset ratio is given in Appendix to Chapter 1, section 1.5.1).

The second paper, entitled '*Foreign currency exposure*', develops the basis upon which the Bank measures and monitors risk to which deposit taking institutions are exposed due to exchange rates movements. The paper sets out the general guidelines for the exposure of banks to foreign exchange risk, and the requirements of information by the Bank for the assessment of the method applied for each institution for measurement of its exposure. Appendix to Chapter 1, section 1.5.2 lists the elements required to be measured and the elements required to be reported and monitored. The third consultative paper, entitled '*The measurement of liquidity*', serves as the basis for assessing the adequacy of liquidity of all institutions covered by the Banking Act.[10] Banks must be able to meet their obligations when they fall due or when they are called. This can be done by holding cash and/or liquid assets and by arranging an appropriate profile of maturing assets. Appendix to Chapter 1, section 1.5.3 provides a description of the liquidity measures.

[9] There are three types of risk:
(1.) credit risk: claims on others not redeemable at the due date at their full book value,
(2.) investment risk -marketable: claims on others may depreciated below their book value, and
(3.) forced sale risk: actual and additional losses sustained because of the need to make ultimately sales of assets may yield less than their quoted value.

[10] Annual reports by the Bank of England, 1979-1982.

During the mid-1980s the sharp growth of off-balance activities and the increasing participation of banks in security markets led the Bank to re-assess its policies by issuing a set of notices. In particular, as increasing number of banks started marketing currency options the Bank issued in April 1984 a memorandum circulated to banks, which states the treatment of banks' foreign currency options in order to measure for the purpose of measuring their foreign exchange exposure. According to the memorandum, exposures arising by exchange rate movements between the two currencies involved must be contained within existing foreign exchange guidelines. The Bank of England considered the hedging technique utilised by each bank in order to validate its appropriateness. If the measure were regarded satisfactory, the Bank would permit the use its own formulae in measuring its exposure on currency options. Otherwise the Bank would take into account 'the potential effect of the exercise of option rights held by the customers of the bank on its open position in individual currencies, and make no allowances either for the likelihood of the option not being exercised or for any options taken by the bank to hedge its position'.[11]

Further to the above, the Bank expressed concerns about the growth of off-balance sheet risks, which were not captured in the measurement of capital. As a result, in April 1985, it introduced a provisional risk asset ratio weighting for underwriting obligations arising out of note of issuance facilities[12] and revolving underwriting facilities. It also announced a review of the whole range of off-balance sheet risks including more traditional contingent liabilities. This was released in conjunction with the *'Statistical notice to monetary sector institutions'*, which stated that a bank's responsibility as an underwriter of a note issuance facility/revolving underwriting facility would be treated as a contingent liability.

The year 1985 was also marked by progress in the development of Banking Supervision in the UK. Specifically, a committee was established to contemplate any changes deemed necessary in the present supervisory system. The committee report,

[11] Annual report by the Bank of England 1984/85, p.p. 8.

[12] These facilities are defined as 'arrangements which enable a borrower to raise funds through the issue of short-term paper, where the availability of funds is in effect guaranteed by a bank or a group of banks underwriting the issue of paper by the borrower. These includes facilities arranged by both bank and non-bank borrowers, where the paper is issued in the form of certificates of deposits or promissory notes. - 'Off-Balance Sheet Risks: Note Issuance Facilities/Revolving Underwriting Facilities', BSD/1985/2, April 1985.

entitled '*The Leigh-Pemberton Report*', was published by the Government in July 1985. The report made the following recommendations:

1. Replacement of classification of recognised banks and licensed deposit-takers under the Banking Act, by a single category of authorisation.
2. Establishment of an operation providing the means for regular dialogue between supervisors and banks' auditors.
3. Concentration of risk in banks' assets must be subjected to more strict controls.
4. Banks' management control systems should be monitor more closely.
5. Increase of staff within the Banking Supervision Division of the Bank.

The recommendations suggested by the report were accepted by the government, which in turn issued in December 1985 a '*White Paper*' on banking supervision to set out the framework for a new Banking Act on Banking Supervision. One of the major additional proposals included in the paper concerned the introduction of a new Board of Banking Supervision.

The proposals of the report and the White Paper did not imply a fundamental change in banking supervision in the UK. Their intention, was to identify a number of weaknesses within the system of supervision, which had become apparent during the implementation of the regulatory framework. Nevertheless, although the reform proposals did not alter the substance of the authorisation requirements of the Banking Act 1979, the government chose to introduce an entirely new bill. Main innovations of the Banking Bill (hereafter Banking Act 1987) were the following:

1. The establishment of the Board of Banking Supervision.
2. The discarding of the two-tier system,[13] which was replaced by a unified system with a single category of authorised institutions.
3. Replacing the provisions on regulations on banking description and names. UK-incorporated authorised institutions were permitted to use a banking name if their paid-up share capital and/or undistributable reserves amount to at least £5 million. Authorised institutions from overseas were entitle to use the names under which they carried business in their country of origin. All authorised institutions were able to use banking descriptions.

[13] The two-tier system had the objective to signal differentiation between the authorised institutions recognised by the Banking Act 1979, in recognised banks and licensed institutions. However, it led to

4. A more detailed description of the authorisation criteria.
5. The Bank had given the power to object to take-overs and to changes in the control of deposit-taking institutions. An advance notice should be given to the Bank if a person is proposing of becoming a shareholder-controller of a UK-incorporated institution. The same stands for existing shareholders that required to take their shareholding over 50% or 70%.
6. The introduction of new requirements on large exposures. Institutions whose primary place of business is UK, should report to the Bank for exposures exceeding 10% of their capital and to give prior notice of proposed transactions that would expose them to risk of loosing over 25% of their capital.
7. In cases where further information of regulatory nature would be requires by the Bank, bank auditors would be released from their duty of confidentiality to their client institution.
8. False provision of information to the Bank would result in criminalisation.
9. The Bank's power to require information and undertake investigations in the affairs of deposit-taking institutions was strengthened.
10. All authorised institutions were required to notify the Bank for changes in directors, controllers and mangers. The Banks had the right to change these requirements for institutions whose principal place of business was outside the UK.

The Banking Bill received Royal Assent on 15 May 1987. The major changes that took place with the Banking Act 1987 were, first, the abolishment of the two-tier status system and its replacement with a single status of authorised institutions. Second, the establishment of the Board of Banking Supervision on a statutory basis, and, third, the significantly increased role of auditors in monitoring and reporting.

At the same year developments in international policy took place as well. In December 1987, the Basle Supervisors Committee issued a paper with the proposals for the convergence of capital adequacy, with three basic principals:

1. There should exist a common definition of capital, where its primary (tier 1 or core capital) value should be equity capital and reserves arising from retained earnings,

confusion, regarding the way related criteria of function and status, were applied. Further, there was no clear distinction in the use of banking names and descriptions between the institutions in the two-tier.

allowing also for a wider range of instruments to be included within a secondary category of capital (tier 2 or supplementary capital).

2. There should exist a common system of risk weights applied to each balance sheet and off-balance sheet items, so as to reflect their credit value.
3. A standard minimum level of capital must exist, which banks would have to hold against their risk-adjusted assets and off-balance sheet business.

The aforementioned proposals have limited discretion to national supervisors in their application of the framework. In the UK, the Bank, where there was a national discretion within the framework, hold to its existing policies, due to the fact that major differences existed between the Basle framework and the domestic system. In July 1988, the Committee issued a new paper[14], setting out a common risk-based system for the measurement of capital, with a required minimum level of 8% for all international banks.[15]

Furthermore, during the same period the European Commission was working on three new Directives on the view of full harmonisation of national regulations towards European integration. The first draft policy paper was on the establishment of '*Second Banking Co-ordination Directive*'. The general idea was the liberalisation of capital movements and other Community instruments in the banking field, generated from the desire to eliminate barriers of establishment in the banking sector. Proposals were based on a single banking license, which would enable credit institutions incorporated in a Member State, to be equally recognised through the EU by virtue of their home country authorisation. With the development of such a license, home country supervisors of the Member State would be accountable for the authorisation and overall supervision of the Community-wide branch operation of their credit institutions. On the other hand, host authorities would be responsible for the

[14] Basle Committee, "International Convergence of Capital Measurement and Capital Standards - The Basle Accord", July 1988.

[15] Further to the above the Basle Committee issued the '*Basle Statement of Principles on Money Laundering*'. In the UK, this Statement was circulated to all authorised deposit takers during January 1989, according to which all UK banking institutions should demonstrate that their policies were consistent with the Statement of Principles. At that time the Bank believed that the Statement reflected the existing best banking practice in the UK. However, in light of increasing international concern about on laundering, the Bank on 10 November of the same year sent a letter to all authorised institutions, to comply with the Basle Statement of Principles. Further, banks had to take up policies in accordance to the statutory provisions relating to the reporting of suspicious transactions to law enforcement authorities. Banks were also notified that failure to install or maintain a system adequate to cope with money laundering could result in revocation of authorisation.

implementation of measures related to monetary policy. The second draft policy paper, called *'Own Fund Directive'*, concerned the definition of capital for prudential supervision purposes. The Directive imposed boundaries to the kind of resources Member States should allow their banks to enumerate as capital. Finally, the *'Solvency Ratio Directive'* introduced a uniform risk weighting system for all EU credit institutions, together with a minimum requirement level for banks' risk asset ratio, set at 8%. The proposals of this Directive were in line with those set out by the Basle Committee.

Following the proposals of the Basle Committee, the Bank of England published a paper in October 1988[16] stating how the proposed framework would be applied to all UK incorporated institutions. Implementation of the Basle Convergence Agreement in the UK came into effect at the end of December 1989. Another paper for capital adequacy purposes was issued in October 1988. The *'Supervisory treatment of the ECU Treasury Bills'* circulated to authorised institutions, according to which, in measuring capital adequacy the Banking Supervision Division, the same treatment to Treasury Bills denominated in ECUs would apply as to those denominated in sterling.

Turning back to domestic policies, UK banks showed a growing interest in recent years in developing schemes for transferring risks associated with banks' loan assets through selling, securitising or sub-participating. Concerns of supervising authorities resulted in the issue of a policy notice in February 1989, entitled *'Loan transfer and securitisation'*. The Bank issued the relevant paper after consulting market participants on the best method by which sale of loan assets can be achieved, together with legal implications and the effects in terms of risk. The policy notice covers the sale of single loans and packaging, securitisation, and sale of loan pools. It also covers the transfer of risk under sub-participation agreements. This policy outlines the conditions that banks must fulfil to eliminate risk when selling an asset. Accordingly, this asset is allowed to be removed from the balance sheet for capital adequacy purposes. Appendix to Chapter 1, section 1.5.4 sets out the objectives of the policy as well as the methods of transfer.

Another policy development that took place in 1989 was the adoption of three EU banking directives. According to the Bank of England, the EC Directives set up the framework for completion of the internal market in banking. The first one is the '*Own Fund Directive'*, adopted in April. This technical Directive determines the items that should be included in the calculation of a bank's capital ('own funds'). It differentiates capital into two categories. The first category, consists of all items that have attributes of capital, such as share capital, accumulated reserves and general provisions (tier 1 or core capital), while in the second category are items of lower quality which can provide protection against insolvency, but cannot guarantee continuing survival of the institution. This includes evaluation reserves, value adjustments and subordinated debt (tier 2 or supplementary capital). The *Second 'Banking Co-ordination Directive'*, was adopted in December and attempted to eliminate the remaining intra-EU barriers to freedom of establishing the banking sector. It permits cross-border business within the EU without need for further authorisation. The main aim of this legislation is to harmonise laws and rules for credit institutions so that they can set up and operate freely across the EU, subject to adequate supervision. It further aims to increase competitions and to remove barriers to banking throughout the EU by deregulating the requirements for branches level of endowment capital. The last one, the '*Solvency Ratio Directive'*, was adopted in December, and its purpose was to incorporate the definition of capital from the Own Funds Directive into a minimum harmonised risk asset ratio of 8% met by all credit institutions. Capital adequacy requirements established by the Directive were in line with the Bank for International Settlements proposals (BIS). However, the BIS capital guidelines were not legally enforceable, as they were only recommendations for international banks. On the other hand, capital adequacy requirements of the Directive were incorporated into EU and member country law, and thus were in fact legally enforceable. Appendix to Chapter 1, section 1.5.5 presents a more detailed analysis for the three aforementioned EU Directives.

In the next decade, banking industry and supervision faced, both in the UK and internationally, significant changes as a result of the creation of the Single European Market. The Bank in December 1990 published a notice implementing the EU

[16] Bank of England, *"Implementation of the Basle Convergence Agreement in the UK"*, BSD/1988/3,

'Solvency Ratio Directive'[17], with amendments to this paper made in November 1992 and March 1995[18], imposing a minimum capital ratio of 8% from 1 January 1993. At the same time, another notice was circulated to authorised institutions establishing the implementation of the *'Own Funds Directive'*[19], with amendments made in January and August 1992[20]. Following this notice, the policy covering the implementation of the Basle Convergence Agreement in the UK was withdrawn.

In 1993, the advent of the Single European Market in Banking required from all Member States minimum standards of supervision. Following that, three EU banking Directives were implemented during 1993 in the UK. The first one is the *'Second Banking Co-ordination Directive'*, which according to the Bank of England is the most important piece of legislation in the establishment of the EU Single Market in banking services. As already mentioned according to this Directive, a bank incorporated in any Member State is permitted to branch or provide services throughout the EU without further authorisation. Legal responsibilities for most aspects of the supervision for EU banks are placed with the home country, while the host country also retains power to impose measures justified on the grounds of general good. [21]

The second Directive is the *'Directive on the Consolidated Supervision of Credit Institutions'*[22]; its purpose was to ensure that home country supervision takes into account total financial business conducted throughout a banking group. In light of that, in March 1993, the Bank published a policy notice notifying authorised institutions on the immediate implementation of the Directive, and all the relevant requirements. Upon the implementation of the Directive, supervision of authorised

October 1988.

[17] Bank of England, *'Implementation in the United Kingdom of the Solvency Ratio Directive'*, BSD/1990/3, December 1990.

[18] Bank of England, *'Implementation in the United Kingdom of the Solvency Ratio Directive'*, BSD/1992/6, November 1992 and BSD/1995/1, March 1995 (amendments to the 1990 paper).

[19] Bank of England, *'Implementation in the United Kingdom of the Directive of Own Funds of Credit Institutions'*, BSD/1990/2, December 1990.

[20] Bank of England, *'Implementation in the United Kingdom of the Directive of Own Funds of Credit Institutions'*, BSD/1992/1, January 1992 (amendment to the 1990 paper), and *'Verification of interim profits in the context of the Own Funds Directive'*, BSD/1992/5, August 1992.

[21] The implementation of the Second Banking Directive, came through the 'Statement of Principles - Banking Act 1987: The Banking Co-ordination (Second Council Directive) Regulation 1992'.

[22] Bank of England, *'Implementation in the UK of the Directive on the Consolidated Supervision of the Credit Institutions'*, BSD/1993/1.

institutions must be conducted on consolidated basis (preparation of consolidated returns covering a group, or part of a group), whenever such institutions are members of a wider group. Moreover, the new notice extended consolidated supervision to include the authorised institution's parent and financial subsidiaries of parent. Furthermore, banks are required to have systems and controls, which are sufficient for the production of data and information on a consolidated basis. Also, authorised institutions are required to submit, at least twice each year, consolidated returns covering capital adequacy and large exposure.

The last Directive is the *'Directive of Monitoring and Control of Large Exposures of Credit Institutions'* implemented in October 1993 with the publication of a policy notice[23] by the Bank. The aim of this Directive was to increase the spread of risks incurred by banks, in order to prevent that default by one client jeopardises the existence of an institution, and the financial system in general. The requirements of the Directive were in line with the Bank's existing large exposure policy, apart from few changes (see Bank of England, *'Implementation in the UK of the Directive on the Monitoring and Control of Large Exposures of Credit Institutions'*, BSD/1993/2).

Moreover, during the beginning of the 1990s banking trading activities continued to develop. More and more institutions throughout EU undertook security-trading activities. Existing requirements up to that time did not address directly market risks faced by banks engaged in such activities, as the solvency ratio captures only credit risk, and interest-rate and currency risks from derivative instruments. In an attempt to accomplish the inclusion of market risks faced by banks that incorporate security-trading activities in the measurement of capital adequacy, and to ensure consistency of treatment between credit institutions and securities and investment firms, the *'Capital Adequacy Directive'* was adopted in 1993. According to the Directive, uniform requirements with respect to the minimum capital cover of trading activities should be applied on both types of institutions (see Appendix to Chapter 1, section 1.5.5). From January 1st 1995 UK-incorporated banks applied the provisions of the Directive to their market risk, and started building the necessary reporting systems. During the same year they formulated trading-book policy statements in detail consultation with the Bank. Furthermore, to allow banks to develop systems for compliance with the Directive by the target date, in December 1994 the Bank issued a consultative paper outlining the proposed approach. Actual implementation of the Directive in the UK came on April 1995 with the publication of a policy notice.[24] A second policy notice was issued in December 1995 to clarify areas of the new policy where necessary.[25]

Today, the responsibility of banking supervision has passed to the Financial Services Authority (FSA). The British Government, in an attempt to reform financial regulation, announced in May 1997 that it would create a single regulator for all financial firm and markets by merging the regulatory responsibilities of nine bodies:

1. Securities and Investment Board (SIB).
2. Supervision and Surveillance Division of the Bank of England.
3. Investment Management Regulatory Organisation (IMRO).
4. Personal Investment Authority (PIA).
5. Securities and Future Authorities (SFA).
6. Insurance Directorate of the Department of Trade and Industry.
7. Building Societies Commission (BSC).
8. Friendly Societies Commission (FSC).
9. Registry Friendly Societies (RFS).

During October 1998 the FSA was introduced, and in June 1988 the Bank of England Act 1998 formally transfers responsibility for banking and wholesale market supervision to the FSA. The FSA starts supply regulatory and other services in January 1999.[26]

We conclude this section with a summary of the requirements from the Bank for any authorised institutions as listed in its published paper "A stable financial system": '*To become authorised, and to remain authorised, a bank must have adequate capital, and must make appropriate provisions against possible bad debts. The aim is to ensure that there are sufficient resources available to the bank to absorb losses*

[23] Bank of England, '*Implementation in the UK of the Directive on the Monitoring and Control of Large Exposures of Credit Institutions*', BSD/1993/2.
[24] Bank of England, '*Implementation in the UK of the Capital Adequacy Directive*', S&S/1995/2, April 1995.
[25] Bank of England, '*Implementation in the UK of the Capital Adequacy Directive*', (amendments to S&S/1995/2), S&S/1995/4, December 1995
[26] Financial Services Authority, Summary Annual Report, 1998/1999.

without placing depositors' money at risk. The bank must also have enough ready cash, or liquidity, to meet likely withdrawals. The quality of management is very important: directors and senior managers must be honest and competent - fit and proper, in the language of the Act. So must those who control banks - major shareholders, for example. There must be adequate internal systems and controls to enable management to assess their risks properly and to ensure that prudent banking procedures are observed.'[27]

Table 1-2 presents all policy and practice notices issued by the Bank of England for the period 1981 to 1995.

[27] Bank of England, *'A stable financial system'*, 1997.

Table 1-2: List of policy and practice for the period 1981-1995.

Title	Date of Issue
Foreign currency exposure	April 1981
Measurement of liquidity	July 1982
Connected lending; accounts ; large exposures; fraudulent invitations; floating charges (BSD/1983/1)	April 1983
Foreign currency options	April 1984
Note issuance facilities/revolving underwriting facilities (BSD/1985/2)	April 1985
Statistical notice to monetary sector institutions (released in conjunction with BSD/1985/2)	April 1985
Large exposures in relation to mergers and acquisitions (BSD/1986/1)	February 1986
Subordinated loan capital (BSD/1986/2)	March 1986
Statistical notice to monetary sector institutions	June 1986
Large underwriting exposures(BSD/1987/1.1)	February 1988
Advertising for deposits	April 1988
Supervisory treatment of ECU Treasury bills (BSD/1988/2)	October 1988
Letter to authorised institutions concerning money laundering	January 1989
Loan transfer and securitisation (BSD/1989/1)	February 1989
Further letter to authorised institutions concerning money laundering	November 1989
Letter to authorised institutions concerning advertising of interest-bearing accounts	December 1990
Letter to authorised institutions concerning guidance notes issued by the Joint money Laundering Working Group	December 1990
Code of conduct for the advertising of interest-bearing accounts	December 1990
Implementation in the UK of the Directive of Own Funds of Credit Institutions (BSD/1990/2)	December 1990
Implementation in the UK of the Solvency Ratio Directive	December

(BSD/1990/3)	1990
Statistical notice to reporting banks on capital adequacy treatment of deferred tax assets	December 1990
Implementation in the UK of the Directive on Own funds of Credit Institutions (BSD/1992/1)	January 1992
Loan transfer and securitisation (BSD/1992/3) (amendment to 1989 paper)	April 1992
Verification of interim profits in the context of Own Funds Directive (BSD/1992/5).	August 1992
Implementation in the UK of the Solvency Ratio Directive (BSD/1992/6) (amendment to the 1990 paper)	November 1992
Letter to authorised institutions concerning debt provisioning	February 1993
Implementation in the UK of the Directive on the consolidated Supervision of Credit Institutions (BSD/1993/1)	March 1993
Statement of principles(Banking Act 1987 section 16; The Banking C0-ordination (Second Council Directive) Regulation 1992 Schedule 7	May 1993
Implementation in the UK of the Directive on the Monitoring and Control of Large Exposures of Credit Institutions (BSD/1993/2)	October 1993
On balance sheet netting and cash collateral (BSD/1993/3)	December 1993
Subordinated loan capital issued by UK-incorporated authorised institutions (BSD/1994/3)	May 1994
Implementation in the UK of the Solvency Ratio Directive (S&S/1995/1) (further amendment to the 1990 paper)	March 1995
Implementation in the UK of the Capital Adequacy Directive (S&S/1995/2)	April 1995
Implementation in the UK of the Capital Adequacy Directive (amendments to S&S/1995/2) (S&S/1995/3)	December 1995

Notes: (1.) BSD: Banking Supervision Division.
(2.) S&S: Supervision and Surveillance.

1.4 Conclusions.

The objective of this chapter was to identify economic and market trends, as well as regulatory alterations, that stimulated changes in the UK banking industry. Changing economic forces during the last two decades generated the need for new financial products and services. Changes in the banks' structure and strategies led to new kinds of financial technology, whereas banking became more international and banks expanded into foreign banking systems. With internalisation of financial services and deregulation in the EU, UK banking industry became more open to equal competition. At the same time there has been a re-regulation trend to capture the new exposures of banks in the face of the new environment.

The major market trends in the UK banking industry during the 1980s and mid 1990s were, first, the improvements in technology that introduced new facilities and new services to customers. Credit cards and cash dispenser, as well as credit transfer and direct debit facilities became readily available in early 1980s, while during 1990s information technology considered as a major part of the improvement process in banks' efficiency. Second, the number of off-balance sheet activities, product innovation, and banks' movements into securities trading increased considerably. In attempting to reduce their risk exposures UK banks increased their involvement in domestic and international securities, while they also tried to increase the liquidity of their portfolios with securitisation. Third, following the international recession during the late 1970s there was a large bad -debt provisions that increased the number of countries experiencing debt-servicing difficulties. Increases in the bad debt provision came forth again during the end of 1980s. Due to worsening economic conditions there were signs of increased difficulties in property, retail and service sector. In the light of that, banks added substantial new provisions, sometimes up to 80%. The high levels of bad debts provisions led to sharp falls in profits. Fourth, there was a liberalisation of cross-border trade of banking and financial services, and the right to establishment of EU financial institutions in other EU member countries.

As far as regulations are concerned, the most important trends are, first, the Banking Act 1979 and its legislative renovation of 1987 with the Banking Act 1987, and, second, the national implementation of the European Banking Directives through secondary legislation and policy notices issued by the Bank of England. The two Banking Acts, emerged from the need to protect investors and the economy, from extensive failures among banks. The national implementation of the EU Banking Directives came from the tendency towards fully harmonised national regulations in view of European integration. According to the Bank the most important regulation in the establishment of the EC Single Market in banking is the Second Banking Directive. Current banking regulations in the UK aim at an optimal trade-off between motivations for quality and competitiveness of bank services, and soundness and stability of the banking industry.

Given all these changes over the last two decades, it is of interest to investigate cost economies and efficiency in the UK banking industry. As deregulation induced substantially changes, and as technology and wider market developments release new competitive pressures, banks are required work at high efficiency levels, and adopt the most cost-efficient size and product mix. Moreover, given the changes in the structure of the banking industry, a main policy issue arises: which type of banking structure serves best the public in terms of both cost and availability of banking services. The current study, taking into the market and regulatory alterations that constitute the major forces of change in the UK banking industry, concentrates on the estimating the performance of UK banks for the period 1980 to 1995, having as a starting point the current chapter, which provides an insight to all of the changes in the industry following the introduction of the Banking Act of 1979 until the Bank of England Act of 1998.

1.5 Appendix to Chapter 1

1.5.1 The measurement of capital

There are two measures for assessing the adequacy of capital: the gearing ratio and risk asset ratio. The first one is the ratio of current liabilities to capital resources, excluding that part of capital that is utilised to finance infrastructure and other non-banking activities. The second one refers to the risk of losses generated with the use of the assets of the institution to the capital that is available to finance such losses. For the calculation if gearing and risk asset ratios the following items of the balance sheet are employed:

Table 1-3 : Composition of gearing and risk asset ratios[28]

Requirements	**Gearing ratio**	**Risk asset ratio**
Capital base	Share capital Loan capital Minority interest Reserves General Provisions	Share capital Loan capital Minority interest Reserves General Provisions
Deduction from the capital base (adjustments)	Investment in subsidiaries and associates Goodwill Equipment Premises Other fixed assets	Investment in subsidiaries and associates Goodwill Equipment Other fixed assets
Adjusted capital base	Deposit and other non-capital liabilities Gearing ratio = 3:4	Adjusted total of risk assets[29] Risk asset ratio = 3 as percentage of 6

[28] Source: Bank of England Quarterly Bulletin, *"The measurement of capital"*, September 1980, p.p. 329.

[29] The classification of assets and the risk weights attached to each of them are given by the Bank-see *"The measurement of capital"*, Bank of England Quarterly Bulletin: September 1980, p.p. 329.

1.5.2 Foreign currency exposure[30]

The paper states the basis on which the Bank would measure, monitor and discuss with deposit taking institutions. In accordance with the paper the following list contains the elements to be measured:

1. The Bank regards exposures generated from any uncovered foreign currency position in any currency, as well as the aggregate net position in all currencies.
2. Future flows of income and expenses not yet received could give rise to an exposure. If those are determinable, then they should be measured.
3. A distinction is required between daily banking operations from those exposures that are long-term in nature.
4. The Bank would consider as long-term exposures those that arise from banks' fixed and long-term assets and liabilities, including items such as loan capital, premises and investments in subsidiaries and associates.

The following list contains elements to reported and monitor:

(A) UK-incorporated institutions:

1. There are no formal limits for a bank's foreign currency position, but there would be agreed position guidelines with each institution individually.
2. The guidelines compare both the dealing position (daily banking operation) and the capital of each bank.

(B) Overseas branches of UK banks: the aforementioned requirements are applicable to all the branches of each bank, both in the UK and overseas.

(C) UK branches of foreign banks: for monitoring purposes, the Bank would take into account the branch's own internal controls, those exercised by the head office, and the monitoring arrangements of its own supervisory authority. If the Bank does not consider these satisfactory, it would agree appropriate absolute levels of exposure to serve as guidelines.

(D) Frequency and method of reporting:

1. So as to monitor the banks' position, the Bank requires returns to be made on a monthly basis. These would give the net spot long or short position and the net forward long or short position in each currency at the close of business on the reporting day. The sum of these positions would give the overall net position in each currency, in current terms.

[30] Bank of England Quarterly Bulletin, *"Foreign currency exposures"*, April 1981 p.p. 235-237.

2. The part of the bank's positions which is agreed to regarded as structural (long-term), should be deducted from the overall net position in that currency to give the dealing position.
3. Swaps should be excluded from the dealing positions.
4. When the dealing position is determined, then it should be translated into sterling pounds.

1.5.3 The measurement of liquidity[31].

The measure is based on a cash flow basis approach, taking liabilities and assets in currency together. In some case it can also be applied to liabilities and assets in one or a group of currencies. It is a 'series of accumulating net mismatch position is successive time bands'. The measure combines the following features:

1. Liabilities: Deposits of all types, known firm commitments to make funds available on a particular date, commitments which are not due to be met on a particular date. Contingent liabilities are not included.
2. Assets: Assets are measured by references to their maturity, marketable assets are taken into account in the cases where they can be sold for cash quickly, contractual standby facilities the bank by other banks. Assets know to be of doubtful value are excluded from the measurement.

1.5.4 Loan Transfer and Securitisation. (BSD/1989/1).

'Transfer by novation and assignments duly notified to the borrower, provided that there are no rights of set-off between the borrower, provided that there are no rights of set-off between the borrower and the seller, are regarded as clean transfers and the relevant assets are excluded from the calculation of the seller's risk-asset ration and included in the buyer's. Less favourable treatment is reserved for transfers by silent agreements (without notification to the borrower) and sub-participation (which do not involve the legal transfer of the right and obligations under the original loan but an entirely separate back-to-back non-recourse funding arrangements): the transfers are disregarded and the assets included in the seller's ratio if the Bank is not satisfied that the full risk has effectively passed to the buyer and that the latter has no recourse to

[31] Bank of England Quarterly Bulletin," *The measurement of Liquidity*", July 1982 p.p. 399401.

the seller. With regard to securitisation, the policy warns the banks originating or servicing the securities against the assumption of residual "moral" obligation to support losses incurred by the buyers of the securities and draws attention to the operational risks.'[32]

1.5.5 EU Banking Directives

1.5.5.1 Own Funds Directive

The Directive aims to establish a common definition and classification for the capital base of credit institutions - banks and building societies - as a basis for setting minimum levels of capacity adequacy. The capital is divided into two categories core and supplementary capital. The latter may be included in the calculation of the own funds only in amounts that do not exceed 100% of the former. Member states were required to bring the Directive into force by 1 January 1993. The Bank recognised the following items as core capital:

1. permanent shareholders' equity;
2. disclosed reserves arising from the appropriation of the retained earnings, share premium s and other surplus;
3. published interim retained profits; and
4. minority core-capital interests arising on consolidation;

but requires the deduction from their total of:

1. goodwill and other intangible assets;
2. current year's unpublished losses; and
3. equity issued by the capitalisation of property revaluation reserves.

The following items are recognised as supplementary capital, up to an overall limit of 100% of the core capital:

1. undisclosed reserves and unpublished interim retained profits;
2. reserves arising from the revaluation of fixed assets;
3. general provisions against loss up to 25% of total risk-weighted assets (but not special provisions);

[32], Hadjiemmanuil, C. "*Banking Regulation and the Bank of England*", UK: LLP limited, 1996, p.p. 202.

4. hybrid capital instruments, i.e. perpetual cumulative preferred shares under certain conditions and only up to 50% of the core capital;
5. minority supplementary-capital interests arising on consolidation;
6. equity issued by the capitalisation of property revaluation reserves.

The following items are deducted from the total of core and supplementary capital:

1. investments in unconsolidated subsidiaries and associates;
2. connected lending of capital nature;
3. all holdings of other credit institutions' capital instruments.

1.5.5.2 Solvency Ratio Directive

It sets minimum standards of capital adequacy that the credit institutions have to meet, by establishing a uniform method of assessing the ability of credit institutions to meet credit losses arising from default of their customers. The directive sets out a risk asset ratio to be applied to credit institutions, where the numerator of the ratio is the capital of an institution as defined in the Own Funds Directive. The denominator is the institution's total assets and off-balance sheet liabilities adjusted to reflect differing degrees of risk. It requires the calculation of solvency ratios on a consolidated basis at least twice per year.

1.5.5.3 Second Banking Directive

It aims at the provision of banking services across the Community. The intuition behind was to harmonise laws and rules for credit institutions so that they can set up and operate freely across the Community being subject to adequate supervision. The main provisions of the directive were as follows:

1. minimum capital requirements for banks of 5 million ECUs, with special provisions for smaller banks,
2. provision for monitoring and venting the bodies that have substantial bank shareholding,
3. controls over banks' long-term participation in non-financial companies, and
4. the establishment of a single banking 'passport' to permit activity within the EC, which is based on the 'home country control' and mutual recognition'.

Member states were required to bring the Directive into force by 1 January 1993.

1.5.5.4 Capital Adequacy Directive

It proposes uniform minimum capital cover for trading activities, for credit institutions, securities and investment firms. The requirements are based on the allocation of a credit institution's holdings of financial instruments[33] between the so-called "trading book" and the non-trading, or banking book.

- Only restrictive positions in the financial instruments held for resale or short-term benefit from price changes and exposures due to unsettled transaction, free deliveries and over-the-counter derivative instruments can be included in the trading book. Allocation of items between the two books must take place on the basis of the objective and consistently applied criteria. The regulated institutions must "mark to market".
- All institutions are required to support their trading-book positions and their foreign-exchange exposures, with capital. Under the "building-block" approach appropriate amounts of capital cover are calculated separately as a fraction of the nominal value of the underlying contract or exposure, for the various identifiable risks inherent in the holding of trading positions. The overall capital requirements is found by adding these amounts.
- Common standards are developed for:
  1. specific (related to changes in the price of the particular instrument) and general (related to broad market-wide movements) positions risk arising from holdings of traded debt and equity instruments, including derivatives as well as underwriting risk arising from the underwriting of such instruments,
  2. counterparty and settlement or delivery risks arising from uncompleted securities transactions,
  3. foreign exchange risks.[34]

[33] This involves transferable securities, units in collective investment schemes, money market instruments, futures contracts, forward interest rate agreements, interest rate, currency and equity swaps, or options to acquire or dispose any of the above

[34] Hadjiemmanuil, C. "*Banking Regulation and the Bank of England*", UK: LLP limited, 1996, p.p. 207.

The Directive set as implementation date the 1st of January 1995, with the national provisions becoming effective by 1 January 1996 at the latest.

CHAPTER 2: The Theory of Modern Banking Firm

2.1 Introduction

One of the crucial problems in finance is to match preferences of surplus sector to lend short, with those of the deficit sector to borrow long. Hicks (1939) characterised this as the 'constitutional weakness' of an unintermediated financial market. This 'constitutional weakness' led to the development of financial intermediaries. Financial institutions interpose between the ultimate lender and the ultimate borrower. The lender places funds with a financial intermediary instead of providing the funds to the final borrower, and in return has a deposit or a claim on the financial intermediary. In turn, the financial firm, having received the funds of the ultimate lender, is then capable of on-lending these funds to the ultimate borrower. The latter provides securities held by the financial intermediary.

But why is it necessary for the financial intermediary to stand between the ultimate lender and the ultimate borrower? The answer is twofold. First, in the absence of intermediation the presence of information costs weakens the ability of the potential lender to find the most appropriate borrower. Second, borrowers and lenders may have different liquidity preferences. The divergence between borrower's and lender's preferences regarding contract terms and direct finance costs, information gathering and monitoring of borrowers, and accepting risks associated with asset transformation, created a space for intermediaries to fill.

Financial institutions are divided into two broad categories: bank and non-bank financial firms. The major feature that distinguishes banks from other financial institutions is the creating aspect of their activities [Partington (1989)]. It is suggested that banks can increase the total volume of spending in the economy by their unique ability to add to the stock of existing credit. On the other hand, non-bank financial institutions play basically a transmission role in the economy, by transmitting funds created by the banking system. Accordingly, banks are central to the smooth functioning of the economy with their uniqueness based on the services they provide.

2.2 The rational for banks

The first step towards studying financial intermediation is to examine the reasoning of bank existence. The question 'why do banks exist' is central to the literature. Recent studies on banks emphasise their role in providing liquidity insurance and risk-sharing opportunities to the agents, by transforming illiquid assets into liquid liabilities, [Diamond and Dybving (1983), Freeman (1988)], minimising transaction costs by acting as delegated monitors of investors, [Diamond (1984, 1989,1996), Chant (1987)], and providing payments and portfolio services.

2.2.1 Portfolio Management

Financial intermediaries transform primary securities issued by firms into indirect financial securities desired by final investors. However, by definition, the chain of transaction between the firm and the final investor is longer than in direct finance additional transactions might lead to an increase in transaction costs. Consequently, any proposition that intermediated finance is more advantageous than direct finance must be based on the view that presumed gains from intermediation outweigh increased transaction costs.

Gurley and Shaw (1960) view intermediaries as the missing link between firms and final investors. In their model, firms issue shares and bonds, whereas the final investor desires demand deposit or life insurance policies. Intermediaries come as connection between the two. They issue demand deposits or insurance policies and hold shares and bonds. But why firms do not issue demand deposits and insurance policies and final investors do not hold shares and bonds? And why intermediaries are actually needed in the Gurley-Shaw analysis? The answer is that direct finance does not permit enough diversification of risk, whereas financial intermediaries are able to do that. Moreover, the function of intermediation in the Gurley and Shaw study can be viewed as overcoming frictions due to transaction costs. Firms could issue the demand deposits required by consumers who face consumption-timing risk, but it might be costly for most consumers to lend directly to a diversified array of firms. By definition, small investors prefer to diversify their risk, which multiplies contracts and transaction costs.

2.2.2 Provision of Payment Mechanism

In developed economies, the payment system is an accounting procedure by which transfer of ownership of certain assets is carried out in settlement of debts incurred. Thus another pivotal role of banks is to facilitate payments and keep rack of wealth among individuals. This booking-activity of banks is realised by debiting and crediting accounts. With the introduction of banks and the payment mechanism that they provide, the holding cost that individuals incurred though the use of currency is reduced by the provision of currency storage and dispensing facilities.

2.2.3 Transformation of illiquid assets to liquid liabilities

With the liquidity[35] function, banks transform short-term liabilities preferred by consumers into long term loans desired by firms. In general, banks are able to transform illiquid assets by offering liabilities with different, smoother pattern of returns over time that the illiquid assets offer, and this illiquidity of assets provides the rational for the existence of banks [Diamond and Dybving (1983)].

An argument formalised by Diamond and Dybving (1983) and expanded by Freeman (1988) is that an important activity of banks is to finance illiquid assets with short-term deposits. In the model proposed by Diamond and Dybving (1983) investors join intermediaries to minimise their risk. But as suggested by Freeman, even risk neutral investors could join intermediaries because they receive expected return through reduction of transaction costs.

Also, intermediary contracts prevent inefficient interruption of production. Provision of liquidity improves a competitive market through risk sharing among people who wish to consume at different random times. While an individual investor faces uncertainty about his consumption path, banks can forecast the intertemporal distribution of aggregate consumption. This makes possible for the bank to promise withdrawals on demand while keeping only a fraction of its assets in liquid though

[35] Liquidity is based on a number of asset properties: (1) Marketability which is associated with the ease and speed with which the value of an asset can be realised, (2) Reversibility which refers to the difference in value between the simultaneous acquirement and realisation of assets, (3) Divisibility which is reflected in the smallest unit in which the transaction in the asset concerned is occurred, (4)

less productive reserves. Basically, the bank's service is to reassign the ownership of the bank's short and long run assets according to realised needs.

2.2.4 Acting as delegated monitors of investors

Another function of banks is that of minimising transaction costs by monitoring loans and signalling in the presence of asymmetric information [Diamond (1984, 1989, and 1996), Chant (1987), Battacharya and Thakor (1993)]. '*Private information held by borrower results in contracting problems, and the delegation of screening and monitoring to banks is an efficient allocation mechanism*'.[36]

In particular, before reaching a decision about a project, investors examine it with respect to its potential productivity. But once the investment takes place, other problems arise. Can it be assured that funds will be directed towards the agreed purpose? Will borrowers fulfil their financial obligations once the repayment is due?[37] To this extent, the monitoring and enforcement problems shift away from ultimate lenders to the intermediary. Because the bank now acts as a delegated agent with the responsibility of monitoring and enforcing, becomes the residual claimant of income. Banks borrow from small investors (depositors) using unmonitored debt (deposits) to lend to borrowers whose loans are monitored.

Leland and Pyle (1977) were the first to present an informational framework for financial intermediaries in the context of delegated monitoring. They suggested that financial intermediaries are able to discover the qualities of individual assets and portfolios and then sell the claims diversified portfolios to investors. They claimed that a bank could attain information about borrowers at lower cost than individual borrowers.

Formalising the ideas of Leland and Pyle, Diamond (1984) indicated the value of diversification in reducing monitoring costs. He suggested that the potentiality of

---

Capital certainty which is the extent to which an asset's future value in terms of cash can be foreseen at future date.

[36] Baltensperer and Dermine (1993), p.p. 26.

[37] The cost of ensuring that funds will be used for the agreed purpose is called monitoring cost, while the cost of ensuring whether borrowers are solvent is called enforcement cost.

financial institutions rests on two propositions. First, monitoring and control of a firm involve natural scale economies. A single financial intermediary monitoring and controlling firms is at least as effective as ten thousand shareholders, yet at a lower cost. Second, if a well established portfolio of firms being financed by intermediary exists, then the relations between the financial intermediary and its own financiers will not be significantly affected by moral hazard and asymmetric information. Based on the two propositions, Diamond showed that under certain situations, an economy with intermediation is more incentive-efficient than with out it.

From the above stated, we see that banks have comparative advantage in monitoring. Also the transformation of short-term liabilities (preferred by consumers) into long-term loans (preferred by firms) must be considered as a joint product with delegated monitoring.

2.3 Banking Regulations

A government tries through the regulations to establish an efficient and stable banking system. The rational for banking regulation rests on two characteristics, namely the social value of banks and the possibility of market failure. As stated in the previous section, banks provide liquidity, transform illiquid assets into liquid liabilities, and act as delegated monitors of investors, along with the traditional function of providing payment and portfolio services. The social value associated with these functions exceeds their private value; the interruption of these services would be costly for the economy as the production could be severely affected and assets are liquidated prematurely. In such a case, markets fail to provide the right quantity of services at the right price, and banks should be regulated in order to ensure the adequate provision of these services. Because banks play a central role to the smooth functioning of an economy, extensive failures among banks could seriously obstruct other sectors of the economy.

Furthermore, in an unregulated economy, there is always higher possibility of market failure. The sources of market failure come from the first two functions of banks. Specifically, risk-sharing deposit contracts leave banks vulnerable to panic runs. This is the core of the most coherent rational for deposit insurance, articulated by Diamond and Dybving (1983). Runs can be provoked by panics, encouraging depositors to withdraw their funds. This results in the premature liquidation of assets and the collapse of what would otherwise be a sound bank.[38]

Additionally, there is risk due to asymmetry of information between banks and investors. Banks lend to companies rather than invest in marketable securities and as a result their activities are difficult to monitor. The evaluation of bank risks is very costly because bank activities are difficult to control, as it is difficult for individual investors to collect information about bank investments. Hence, monitoring and evaluation of banks by each depositor separately is not possible. Nevertheless, perfect information is a necessary condition for the existence of competitive markets. Otherwise investors cannot access banks at all times, which creates again the problem of market failure. In turn, if investors do not realise that a single bank failure is not a sign of total collapse of the system, then panic runs may be observed. Therefore, externalities between investors and banks justify regulation.

In determining the scope and the forms of regulation, regulators must bear in mind that the two classes of market failures described above are very different in nature. Firstly, the systematic risk caused by bank runs, where investors face the risk of insolvency, is basically financial in nature. To avoid this risk a central requirement is that banks hold sufficient capital so as to reduce the risk of market failure. On the other hand, the risk arising from asymmetric information reflects the nature and the activities of a firm and its employees. Consequently, regulations are unlikely to play an important role to the extent that they are indirectly related to the nature and the activities of the firm and its employees, since any correction of asymmetric information is based on screening and monitoring of firms and its employees.

[38] Three conditions are necessary to make panics possible in the absence of any regulations; First, banks must satisfy a sequential service constraint, that is withdrawals tenders must be served sequentially till the bank runs out of assets. This creates incentives to run and get the money before other people. Second, investments of the bank cannot be totally liquid. Differently by withdrawing early it is not possible to gain anything. And third, depositors must have high enough degree of risk aversion. Otherwise the optimal risk sharing contract involves a face value lower than the liquidation of bank assets.

2.4 Measuring Bank Output

So far, the intermediary and payment functions were identified as reasons for bank existence and rational for regulating banking industry. It must be admitted though that their activities raise definitional problems of the determination of their production. The literature has exposed considerable difficulties in the definition and measurement of bank input and output, since financial institutions are multi-product firms producing services rather than physical products. Output of primary and secondary industries can be measured in terms of physical quantities or money values deflated by appropriate price indices, whereas output in the form of services usually cannot be measured in terms of physical quantities. In addition, the output and input of financial institutions presents particular difficulties because each bank is a multi-product firm, where many of its services are joint or interdependent, and also banking is subject to government regulation that may affect cost, prices or its level of output [Kinsella (1980)]. As a result, economists disagree on the correct definition of output and input in the banking industry, because each definition embeds a particular set of banking concepts.

To define input and output of banks two main questions should be answered. First, how to choose the appropriate output and input and second, how to measure these variables. As far as the latter is concerned, the solution is constrained by data availability and requires choosing if inputs will be measured by flow or stock variables. In the banking industry it is appropriate to consider some stocks as a proxy of outputs (although this is unusual for other sectors of economic activity), given the assumption that deposits and outstanding loans require a continuous flow of work independent of customer's demand. Furthermore, stock measures can be chosen because of data availability.

Benston, *et al.* (1982) have described the above problem in the following manner: '*One's view of what banks produce depends on one's interest. Economists who are concerned with economy-wide issues tend to view bank's output as dollars of deposits or loans. Monetary economists see banks as producers of money-demand deposits.*

Other sees banks as producing loans, with demand and time deposits being analogous to raw-materials'.[39]

The lack of consensus in the relevant literature leaves the definition of input and output unsettled, and up to now there is no precise definition. Most prior banking studies did not use national accounting measures; instead they tended to adopt either the *production approach* or the *intermediation approach*. According to the *production approach* banks are viewed as firms that are mainly producers of demand deposits, time and savings deposits, and loans account services. Output is measured by the number of account serviced or the number of transactions carried out on each type of product, while input cost is total operating cost incurred in the production of these outputs. Output is treated as a flow, i.e. as the amount of output produced per unit of time, and inflation bias is absent. In this approach the definition of inputs includes capital and labour but precludes interest costs. The production approach is appropriate for studying cost efficiency of banks since it concerns operating costs.

On the other hand, in the *intermediation approach* banks are viewed as intermediators of financial services. They transform and transfer financial resources from units in surplus to units in deficit. The value of loans, investments, and other assets are viewed as appropriate measures of bank's output. Inputs include total production costs, which are labour, capital, operating cost plus interest costs. As far as deposits are concerned, some studies use them as bank input and others as bank output. This approach is merely used for studies where most activities of banks involve are purchasing funds from other financial institutions and large deposits, and turning them into loans and financial investments.

The *intermediation approach* was first used in early cost studies. Alhadeff (1954) and Horvitz (1962) measured bank output as the ratio of loans and investments, to total assets to reflect used capacity to total capacity of the bank. Moreover, average costs were measured by the ratio of total expenses to loans and investments. In this sense, earning assets were used as the measure of bank output. The major disadvantage of this measure is that it excludes other assets; this omission will tend to inflate the unit

[39] Benston et al. (1982).

cost of larger banks. The studies of Schweger and McGee (1961) and Gramley (1962) were not subject to this potential bias because total assets and deposits were used to normalise expense data. However, all these studies used real-valued unweighted indices, which ignored the differential importance of individual bank products, the relative cost of production, and the easiness with which banks can change their product mix. Further the amount of assets and deposits are 'stock' concepts, while production is a 'flow' concept.

To correct some of these problems, studies in late 60's and 70's have used weighted indices to measure banks' output. For example, Greenbaum (1967) used the market value of services to measure output in an attempt to estimate the real social value of banking services. He considered demand and time deposits as inputs, and output was expressed in terms of assets. With the use of linear regression methods he derived a set of average interest rates charged on various categories or earning assets which were used as weights; however, a major problem of this approach was that it ignored the effect of inflation on interest rates.

Meanwhile, Benston (1965) introduced the *production approach*. The advantage of this method is that it allows numbers of accounts and average size of accounts to have differential effects on costs. Further, it meets some of the problems of the intermediation approach by removing the inflation bias and is a flow concept. On the other hand, this approach suffers from lack of a method for weighting the contribution of each service to total output and omits many important items of bank services such as interest costs. Benston, Hanweck and Humphrey (1982) used a new output measure called the Divisia Multilateral Index, which is defined as the weighted numbers of accounts in each activity area. Again, the approach is still vulnerable to criticism due to the omission of interest cost, which constitutes a substantial proportion of bank's total costs. This approach dominated the literature until the beginning of the 1980s. In more recent studies this approach has only be used by studies focusing on the relative efficiency of branches within a particular bank, rather than across banks. Latest studies use the intermediation approach, based on the notion that output should be measured as the dollar value of banks' earning assets.

Berger and Humphrey (1992b) stated that before reaching a decision on which of the above method is to be used, an identification of the most important banking function for the study being undertaken is required. Consequently, they outlined three approaches to this initial identification process.

The first one is the *asset approach*, which is a variant of the intermediation approach. According to this approach, assets (mainly those that produce loans) strictly define outputs. Banks are considered as financial intermediaries between liability holders and bank funds recipients. Loans and other assets are considered as output, whereas deposits and other liabilities are inputs to the intermediation process. However, this approach has one major drawback. For some large banks, that basically purchase their funds from other banks and turn these funds into loans, the asset approach is a satisfactory description. Nonetheless, most banks, especially nowadays, do more than simply purchasing their funds; they provide a series of services to their depositors, which are not counted as output in this approach. For studies of loan cost or studies of bank's profitability *asset approach* is the most appropriate because costs and different methods of raising funds are taken as exogenous variables (assuming the utilisation of a reduced form model). In contrast, studies that consider total banking output need to consider a structural form in which purchased funds are an intermediated output of raising deposits, and the services provided to depositors are a partial payment for these funds.

The *user cost approach* determines whether a financial product is defined as an input or output on the basis of its net contribution to the bank revenue. If the cost of a particular liability is less than the opportunity cost or if the returns of an asset exceed the opportunity cost, then the item is considered as an output, while otherwise it is considered as an input. '*An optimising bank earns exactly its opportunity costs of funds at the margin on each asset and pays exactly its opportunity costs at the margin on every liability. Thus, to the extent that the user cost approach accurately measures marginal financial revenues and opportunity costs, its allocation is largely on the basis of excluded operating costs.*'[40] Nevertheless, there are some problems in measuring financial revenues and marginal opportunity costs, which makes the

[40] Berger and Humphrey (1992a) p.p. 247.

determination of bank output and input subject to significant measurement error and sensitive to changes of data over time. A major difficulty involves the presentation of assets and liabilities in the balance sheets of the banks. Banks frequently use compensated balances or pay below market rates on deposits as a method of charging for bank services. Further, sometimes borrowers required to hold part of their loans as idle demand deposit balances. Consequently, some of the bank's earnings on loans are less than the opportunity cost of funds on deposits. Hence, there is a bias towards finding loans to be inputs or to have a smaller output weight, and towards finding demand deposits to be an output or to have a higher output weight.

Another problem is the adjustment of the opportunity cost for important characteristics of bank assets and liabilities, such as differences in maturity and credit risk. According to theory, the financial return or cost of each category must be adjusted before applying a common opportunity cost. However these adjustments are difficult to be implemented in practice.

Finally, another approach is the *value-added approach.* The difference of this approach from *asset* and *user cost approaches* is that it considers all liabilities and assets categories to have some output characteristics, rather than distinguishing inputs and outputs in a mutually exclusive way. In this approach the identification of activities as inputs and outputs is based on the share of value added, which means that items of balance sheet with substantial share of value are considered important outputs. Under the *value-added approach* the major categories of produced deposits, namely demand, time and savings, and loans are considered as output, whereas purchased funds are considered as inputs because they require very small amounts of labour and capital. Furthermore, government securities and other non-loan investment are considered to be unimportant outputs, because again their value-added requirements are very low. Table 1 gives a summary of input and output measures of latest studies together with the approach used in the determination of these inputs and outputs.

Table 2-1: Input-Output Measures in Bank Cost Literature.

Authors	Inputs	Outputs	Approach
Benston, Hanweck, and Humphrey (1982)	Labour Capital	Demand deposits Time and saving deposits Real estate loans Instalment loans Business Loans	PA
Gilligan and Smirlock (1984)	Deposits	Loans Total Securities	PA
Berger, Hanweck, and Humphrey (1987)	Labour Capital	Demand deposits Time and savings deposits Real estate loans Commercial loans Instalment loans	PA (No of a/cs and average size of a/cs) IA (dollar)
Aly, Grabowski, Pasurka, Ragan (1990)	Labour Capital Loanable funds	Real estate loans Commercial and estate loans Other loans Demand deposits	IA
Farrier and Lovell (1990)	Labour (employees) Occupancy costs and expenditure on furniture and equipment Expenditure on materials	No of demand deposit accounts No of time deposits accounts No of real estate loans No of commercial loans No of instalment loans	PA
Berger and Humphrey (1991)	Labour Physical capital Purchased funds Large certificates of deposit Foreign deposits Other liabilities for borrowed money	Deposits (demand, retail and savings) Loans (Real estate, commercial and industrial, instalment)	IA
Glass and McKillop (1992)	Labour Capital Loanable Funds	Lending output Lending plus non-lending output Total earning assets	IA
Berger and Humphrey (1992)	Labour Physical Capital Purchased Funds	Demand deposits Time and savings deposits Real estate loans Commercial and industrial loans Instalment loans	IA
Nathan and Neave (1992)	Labour Physical Capital	Commercial and industrial loans	IA (1,2 and 5)

	Financial Capital	Other loans Time deposits Demand deposits Securities and investments	VAA (1,2,3, and 4)
Mester (1992)	Labour Physical capital Deposits Borrowed funds	Loans required monitoring New loans originated Loans purchased Average loans sold	IA
Grabowski, Rangan, Rezvanian (1993)	Labour Capital Loanable Funds	Loans (Commercial and industrial, consumers loans) Demand deposits Investment securities	IA
Bauer, Berger and Humphrey (1993)	Labour Physical Capital Purchased funds	Demand Deposits Small time and saving deposits Real estate loans Commercial and industrial loans Instalment loans	IA
English, Grosskopf, Hayes, Yaisawarng (1993)	Interest bearing small deposits Labour Occupancy expenses Purchased or borrowed funds	Loans (real estate, commercial, consumer instalment loans) Investments	AA
Fields, Murphy, and Tirtiroglu (1993)	Labour Capital Credit Total Deposit	Total amount of loans	IA
Allen and Rai (1993)	Labour Capital Loanable funds	Wholesale and retail loans Insurance and real estate loans Investment and trading assets	IA
Noulas, Miller, and Ray (1993)	Labour Capital Deposits All other loanable funds	Real estate loans Consumer loans Commercial and industrial loans Federal funds sold, total securities, and assets held in trading a/cs	IA
Berger, Hanckock, and Humphrey (1993)	Labour Purchased funds	Business and consumer loans Deposits Services (number of branches)	VAA
Favero and Papi (1993)	Labour (employees)	Loans Investments	AA IA

	Capital (fixed assets and premises) Loanable funds	Non-interest income (only AA)	
Berg, Forsund, Hjalmarsson, Suominen (1993)	Labour (man-hours per year) Capital (machinery and equipment)	Loans to other financial institutions Deposits from other financial institutions Number of branches	PA
Mester (1993)	Labour Physical capital Deposits and other borrowed money	Mortgage loans Other loans Securities and other investments	AA
Berger (1993)	Labour Capital Core deposits Purchased Funds	Demand deposit Retail and time saving deposits Real estate loans C&I loans Instalment loans	VAA
McAllister and McManus (1993)	Purchased Funds Saving deposits Fixed assets Labour	Real estate loans Commercial and industrial loans Instalment loans Demand deposits Saving deposits	VAA
Grabowski, Rangan, and Rezvanian (1994)	Labour Physical capital Loanable funds	Real estate loans Commercial and industrial loans Consumer loans Demand deposits Securities	IA
Kaparakis, Miller, and Noulas (1994)	Deposits Labour Physical Capital Purchased Funds	Loans to households Loans secured by real estate Commercial and industrial loans Federal funds sold and total securities	IA
Elysiani and Mehdian (1995)	Labour Capital Saving deposits Demand deposits	Real estate loans Investments Commercial and industrial loans Other loans	IA
Jagtiani, Nathan, and Sick (1995)	Labour Physical capital Financial Capital	Earning assets (loans and investments Deposits (VAA) OBS guarantees OBS foreign currency OBS interest rate Aggregate OBS items	IA VAA

Miller and Noulas (1996)	Total transaction deposits Total non-transaction deposits Total interest expenses Total non-interest expenses	Commercial and industrial loans Consumer loans Real estate loans Investments Total interest income Total non-interest income	IA
Jagtiani and Khanthavit (1996)	Labour Physical Capital Deposits	Produced deposits Loans and Investments OBS Products	IA
Lang and Welzel (1996)	Labour Physical Capital Deposits	Short-term loans to banks Long-term loans to non-banks Interbanking assets Residual output Fees and commissions Revenues from sales of commodities	IA
Mester (1996)	Labour Physical Capital Deposits	Real estate loans Lease financing receivable, agricultural loans, loans to foreign governments, other loans Loans to individuals	IA
McKillop, Glass, and Morikawa (1996)	Labour Capital Loanable funds	Loans and Bills Discounted Cash and due from banks plus call loans Securities plus trading account securities	IA
Berger, and DeYoung (1997)	Labour Physical capital	Commercial loans Consumer loans Real estate loans Fee-based income	AA
Berger, and Mester (1997)	Labour Core deposits Purchased funds	Consumer loans Business loans Securities	IA

Notes: PA denotes Production Approach,
IA denotes Intermediation Approach,
VAA denotes Value-added Approach, and
AA denotes Asset Approach.

2.5 Competitive Issues in Banking

From the beginning of 1970s, world banking experienced marked changes. The expression 'financial revolution' has been coined to characterise the scale and nature of these changes. Many countries adopted structural deregulation policies to liberalise their financial institutions and markets to make them more open to equal competition. Together with structural deregulation trends, there has been a growing conduct re-regulation pattern, as domestic supervisory systems had to adapt to new developments in international banking and finance and to cope with new kinds of risk exposure.

Concentration in financial market increased with a trend towards a smaller number of larger institutions in many leading banking sectors. Furthermore, banking became more open to world developments and banks expanded into foreign banking systems. Also, technology had a significant impact on banking in areas like competition, new products and payment system. But above all, the importance of banking services in economic activity grew substantially. The number of bank branches and the level of employment in the banking sector increased in the major European countries during the 1980s. An EC study indicated that contribution of financial services during the first part of the 1980s amounted to 6.5% of total value-added and accounted for around 3% of total employment within of the member countries.[41] Last but not least, the banking sector grew in size relative to gross national product in almost every European country over the last decade.[42]

Important elements that impact on the new face of the banking industry relate to competition, cost efficiency, and changes in the market structure. First, in the light of recent deregulation and its potential effects on market competition managers and shareholders are interested in knowing how to improve cost efficiency so as to obtain higher profits and increase chances of survival within the industry. Next, customers are interested in knowing the effects of these changes on bank cost, and consequently on price and quality of bank services. Finally, policy makers would like to have

---

[41] European Economy (1988).
[8] Gardener and Molyneux (1994), ch.2.

information concerning the effect of the present regulatory framework on bank cost, competition and market structure of the industry.

2.5.1 Scale and Scope Economies

Scale economies suggest that larger firms produce output at lower average cost than smaller firms [Lewis and Davis (1987)]. A strict definition is in terms of production function. That is, a cost saving resulting from economies of scale exists when the proportional increase in all inputs used in the production leads to a greater than proportional increase in output. Nevertheless, the observation that per unit costs fall as output increases does not necessarily suggest the existence of economies of scale. The reduction in cost per unit of output may be caused by other factors such as technological changes, or improvements in managerial ability (X-efficiency). Hence, in order to examine the effects of scale only on cost, we must separate the effects of other factors [Kolari and Zardakoohi (1987)].

Scale economies are based on the shape of the average cost curve, which illustrates average costs at each level of output. Figure 2-1 displays the long-run average cost (LAC) curve and long-run marginal cost (LMC) curve with a series of short-run average cost (SAC) and short-run (SMC) curves.

The average cost curve shows the average cost per unit at different levels of output, while the marginal cost curve shows the additional cost incurred when producing a very small increment of output. The Long-run curves allow for simultaneous changes in all factors of production, and Short-run cost curves represent cost changes as output increases because of changes in some factors of production. Each set of short-run cost curves represents a different amount of fixed factors of production; e.g. the fixed factors associated with SMC_2 and SAC_2 are greater than those for SMC_1 and SAC_1. The LAC curve is declining at output up to Y_m, where economies of scale (or increasing returns to scale) exist. For level lower than Y_m LMC lies below LAC, whereas for higher levels, LMC lies above the LAC, and diseconomies of scale (decreasing return to scale) exist.

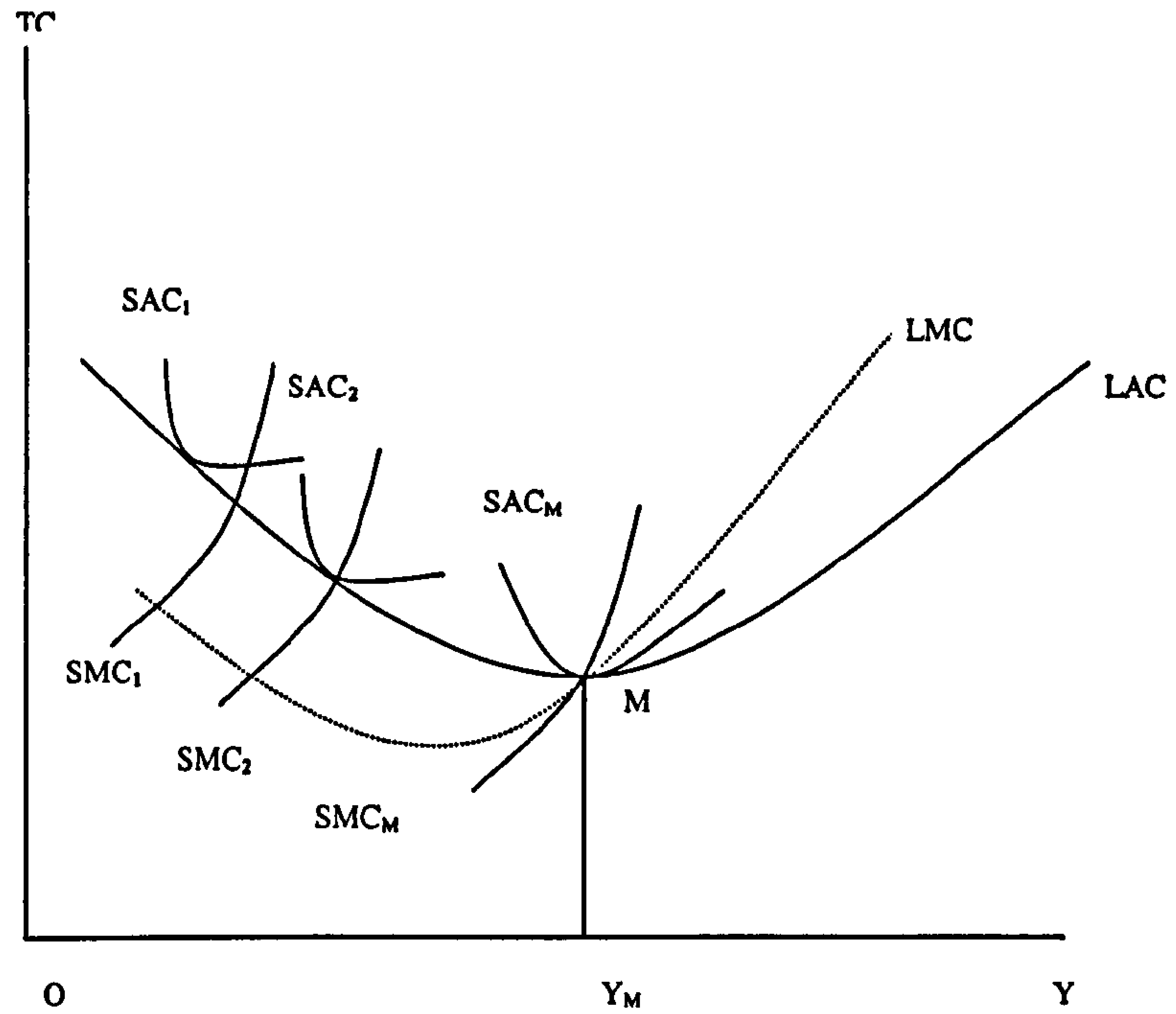

Figure 2-1: Average and marginal cost curves and economies of scale

Figure 2-1 presents a U-shape cost curve. The shape of the *LAC* curve is important because of its implications for scale decisions and of its effect on the potential level of competition in the market. For instance, a flat-bottomed *LAC* curve indicates the existence of constant returns to scale. The number of competitors and the ease of entry will be greater for industries with U-shaped *LAC* curves than with those with L-shaped or continuously downward sloping *LAC* [see Geroski (1991)].

From the viewpoint of a bank, economies of scale exist if an increased size lead to an increased return of shareholders' funds [see Lewis and Davis (1987)]. Certain sources of size economies highlight this advantage. First, by interpreting scale via balance sheet items and based on the assumption that all items in the profit and loss move proportionately to scale, profits will increase more than proportionately because physical inputs and, consequently, operating expenses grow less rapidly than the scale of activities. As a result, the rate of return on shareholders' funds will rise. Second, in the existence of scale economies, there is the possibility to rearrange the asset portfolio to grow faster than interest paid. This implies diversification across liabilities and assets with uncertain maturities as size increases, and operate with a lower ratio of low-yielding reserve assets. Finally, an increase in size might lower the variability in profits. When the income earned on assets follows a stochastic process, a bank can operate with a higher degree of leverage and as a result the rate of return on shareholders' funds will rise for a given degree of risk.

Moreover, economies of scale play a significant role for the regulatory authorities. In an unregulated economy, where the banking industry is characterised by economies of size, larger banks dominate because they would be more efficient. Furthermore, because as competition increases, concentration will increase as well, because cost inefficient banks will no longer be viable. Consequently, banking industry will consist of larger and, probably, fewer institutions.

In terms of product mix, cost reductions can be achieved through economies of scope. Many firms produce more than one product, and in some cases the firm's products are closely linked. Thus, firms are likely to enjoy cost advantages when they produce more than a single product. In strict terms, **economies of scope** are present when joint output of a single firm exceeds potential output of two different firms each producing a single product (assuming equivalent production inputs). On the other hand, if a firm's joint output is less than that of separate firms, then its production process involves diseconomies of scope. The concept of economies of scope can be explained geometrically in Figure 2-2.

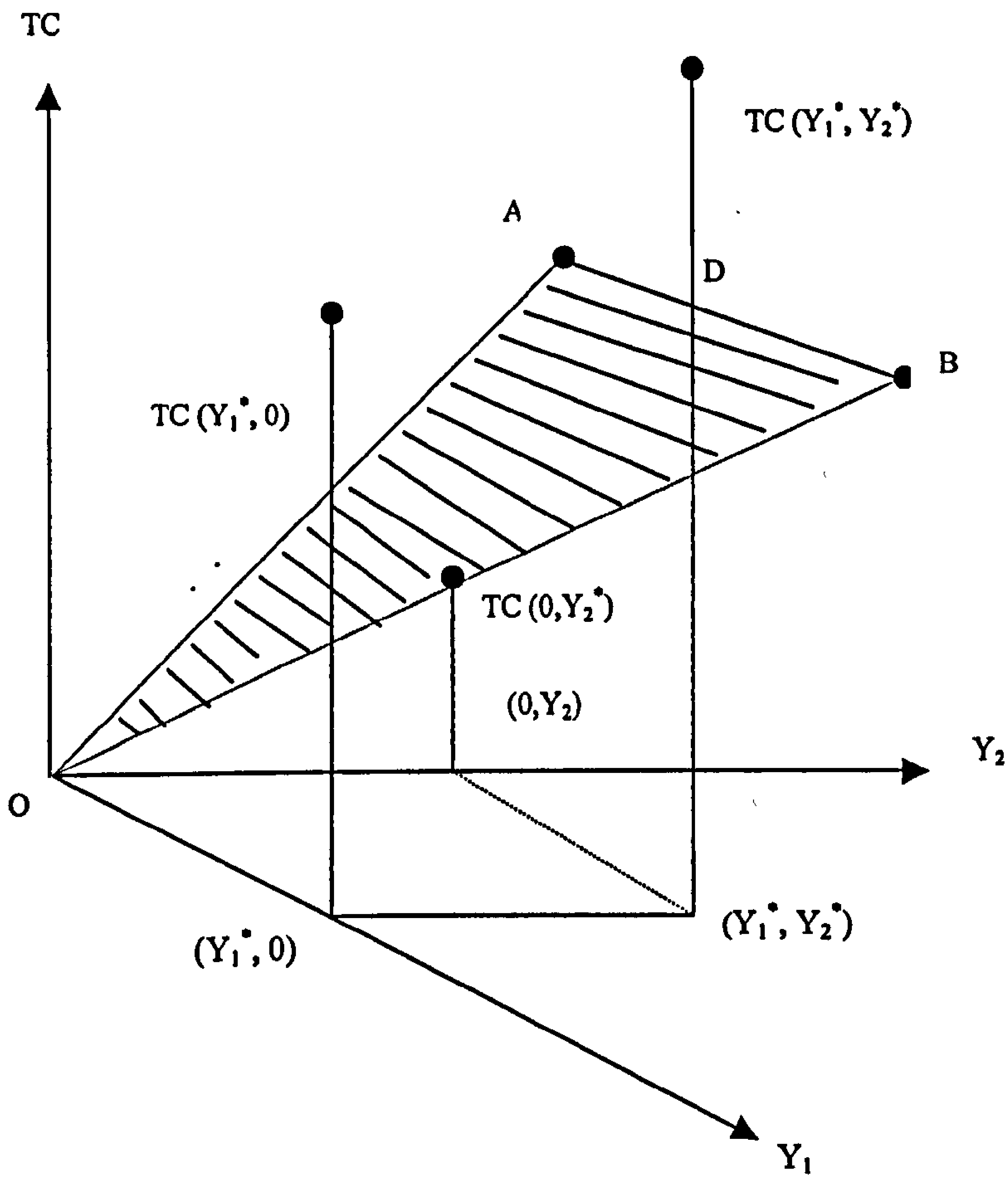

Figure 2-2: The concept of economies of scope (Source: Molyneux *et al.* (1996), p. 144).

Consider two outputs, Y_1 and Y_2, and the separate cost functions $TC\ (Y_1)$ and $TC\ (Y_2)$ respectively. If there exist economies of scope then:

$$TC\,(Y_1,\ Y_2) < TC\,(Y_1) + TC\,(Y_2) \tag{1.1}$$

Figure 1-2 illustrates that the scope economies concept entails a comparison of $TC\ (Y_1^*,\ 0)$ plus $TC\ (0,\ Y_2^*)$. These are the sum of heights in the surface over the corresponding points on the axes, with $TC\ (Y_1^*,\ Y_2^*)$, the height of the cost surface at point $(Y_1^*,\ Y_2^*)$ which is the vector sum of $(Y_1^*,\ 0)$ and $(0,\ Y_2^*)$. If $TC\ (Y_1^*,\ Y_2^*)$ lies below the hyperplane *OAB* which goes through the origin and points $TC\ (Y_1^*,\ 0)$ and $TC\ (0,\ Y_2^*)$, then the conditions for scope economies are attained. Point *D*, in figure 2-1, must equal $TC\ (Y_1^*,\ 0) + TC\ (0,\ Y_2^*)$ since the hyperplane is defined by $TC = \alpha Y_1$

+ βY_2 for some constants α, β. Accordingly, $TC\ (Y_1^*, 0) = \alpha Y_1^*$ and $TC\ (0, Y_2^*) = \beta Y_2^*$ must be less than $\alpha Y_1 + \beta Y_2$ for scope economies to hold [Baumol *et al.* (1988)].

As already mentioned above, economies of scope imply that joint production of two or more products by a single firm is less costly than the sum of their production by two or more firms. Now, banking by definition is a business that entails the joint production of deposits and loans. In addition, the existence of sharable goods, such as bank employees and technology, generates the possibility of their being. The use of such inputs for the production of an output does not rule out the possibility of the same input to be used in the production of other outputs. On the contrary, due to the nature, these inputs can be used, to a certain extent, for the production of several outputs. Furthermore, joint production of different outputs results in the spread of fixed costs and generates information economies by the costless use of information obtained in the context of other services. Last, but not least, product mix may lead to risk reduction through asset diversification across different asset group, which in turn may reduce portfolio and interest rate risk.

Again, the existence of such economies in the banking industry is equally important. From the aforementioned definition, it is clear that economies of scope underlie some of the limitations of specialisation, because specialised banks will be less efficient than multi-services bank. And in an unregulated environment these institutions tend to die and be replaced by non-specialised banks.

2.5.2 X-Efficiency

In evaluating the performance of a firm, the first task is to separate production units that perform well by some standards from those with poor performance. This can be achieved through frontier analysis. In the economic literature **X-efficiency** covers all technical and allocative efficiencies of individual firms, distinct from scale and scope economies. Technical efficiency determines the proportional decrease in input usage, which would have been achieved if the firm operated on the production frontier. While allocative efficiency indicates the proportional decrease in cost if the right input mix had been utilised.

Modern literature on efficiency measurement and on estimation of production frontiers starts with the seminal article by Farrell (1957). Farrell's measures are still generally accepted and widely used until nowadays. Figure 2-3 displays X-efficiency.

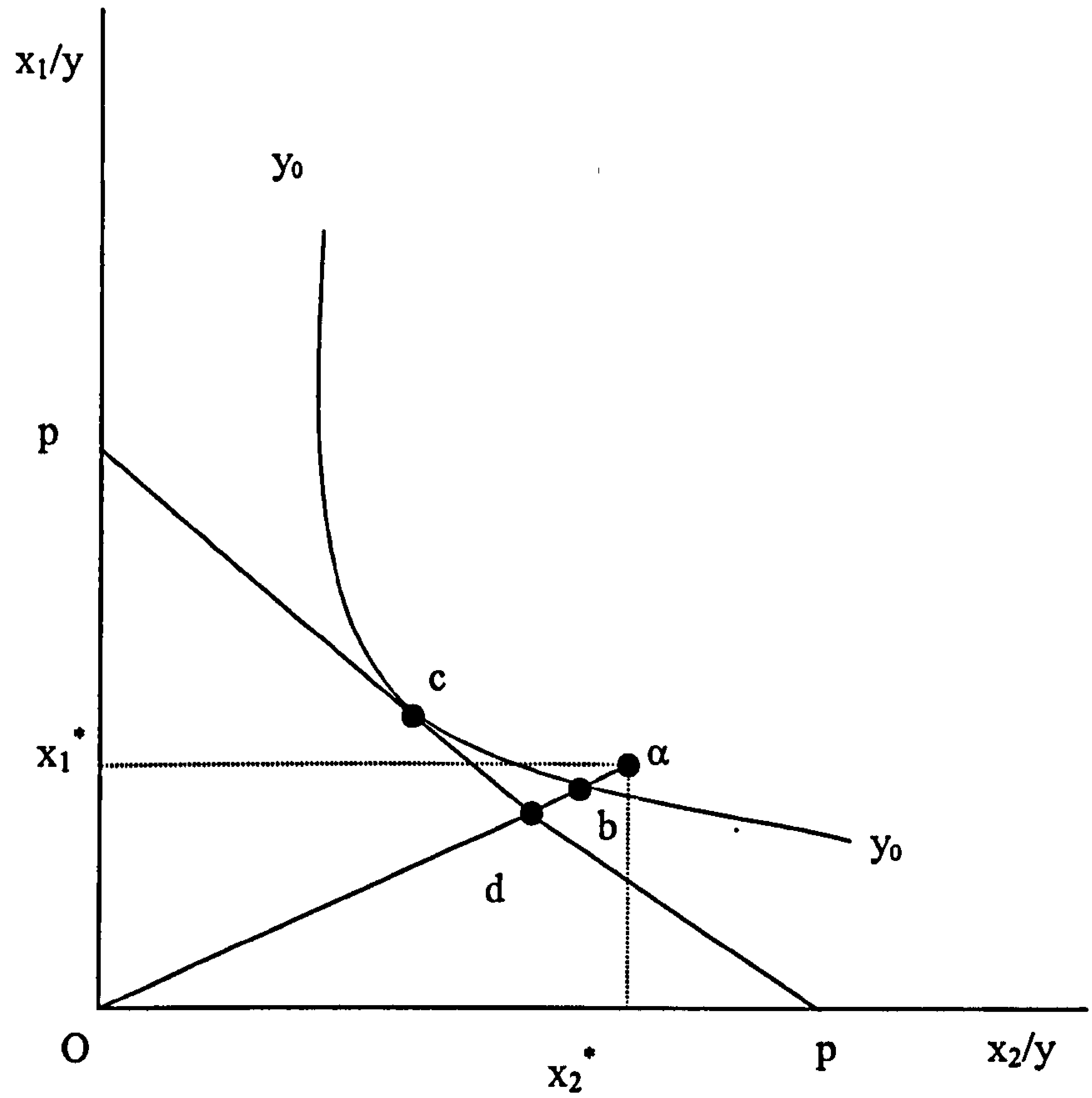

Figure 2-3: Overall, Technical and Allocative Efficiency (Source: Schmidt (1986), p.292).

Assume two inputs (an assumption made for graphical simplicity only) so that the production frontier is $y = f(x_1, x_2)$ where x_1, x_2 are inputs and y is the output. Further assume constant returns to scale so that $f(x_1/y, x_2/y) = 1$; the frontier is characterised by the efficient unit isoquant, illustrated in Figure 2-3 as y_0y_0. The x_1/y and x_2/y are the vertical and horizontal axes respectively. Suppose that the firm uses x_1^* and x_2^* to produce y^* and is denoted as point α in Figure 1-3. The necessary proportion of (x_1^*, x_2^*) for the production of y^* is given by $Ob/O\alpha$, where it denotes the technical efficiency of a firm. Accordingly, $1- Ob/O\alpha$ indicates the technical inefficiency of the above firm. Technical inefficiency (T) measures the proportion by which (x_1^*, x_2^*) could be reduced without reducing output, holding the input ratio x_1/x_2 constant.

Consider now *pp*, which shows the ratio of input prices, with the point *c* as cost minimising. However, since the cost at point *d* is the same as the cost at *c*, allocative efficiency (A) is measured by the ratio *Od/Ob*, and correspondingly the allocative inefficiency of the firm is given by *1 - Od/Ob.*

Total efficiency (X-efficiency) of the firm is defined as $Od/O\alpha$, which measures the proportional reduction in costs which could have been obtained if the firm had been both allocatively and technically efficient. The relationship between overall, technical and allocative efficiency is given by the following [Farell (1957)]:

$$OE = T + A. \tag{1.2}$$

OE measures the possible reduction in the cost from moving from point α (observed cost point) to *c* (the cost minimising point). Total inefficiency can be decomposed as the sum of technical and allocative inefficiencies [Schmidt (1986)].

In general terms, X-inefficiency measured as the deviation from the efficient frontier is the effect of the differences in managerial ability to maximise revenue or minimise cost. The term X-inefficiency has first been formulated by Leibenstein (1966) and covers the excess of actual costs over the minimum attainable level given the current state of technology. Essentially, this kind of efficiency is measured, or modelled, as the distance reflecting a firm's own position in relation to its efficient production (cost frontier). The real advantage of frontier estimation is that it permits objectively determined numerical efficiency values and ranking of firms that is not otherwise available [Berger and Humphrey (1997)].

Turning to the banking sector, the question whether banks are efficient is of interest to managers and shareholders, as well as to customers and regulators. If these institutions become more efficient, then profitability will improve, greater amounts of intermediated funds will be available and better prices and service quality for consumers will be obtained. Thus, banks become more viable, and therefore, the possibility of failure or being subject to take-over, decreases substantially.

2.5.3 Market Structure and Performance

Apart from the internal structure of the firm, performance is also affected by external factors. The broad line of reasoning is that firms conduct pricing and production policies under market rivalry, and hence their performance (profitability and efficiency) depends on market structure.

But which market structure satisfies best firm's performance? From the policy-makers viewpoint, which market structure maximises social welfare? For instance, let us consider two extremes cases, perfect competition and monopoly. Under perfect competition, combined marginal cost equals the market price. By contrast, under monopoly price is greater than cost. The different predictions concerning price and output in long run equilibrium in terms of dead-weight welfare loss are illustrated in Figure 2-4. The critical assumption underlying this model is that cost conditions are the same for a monopolist and for a set of small firms operating under perfect competition.

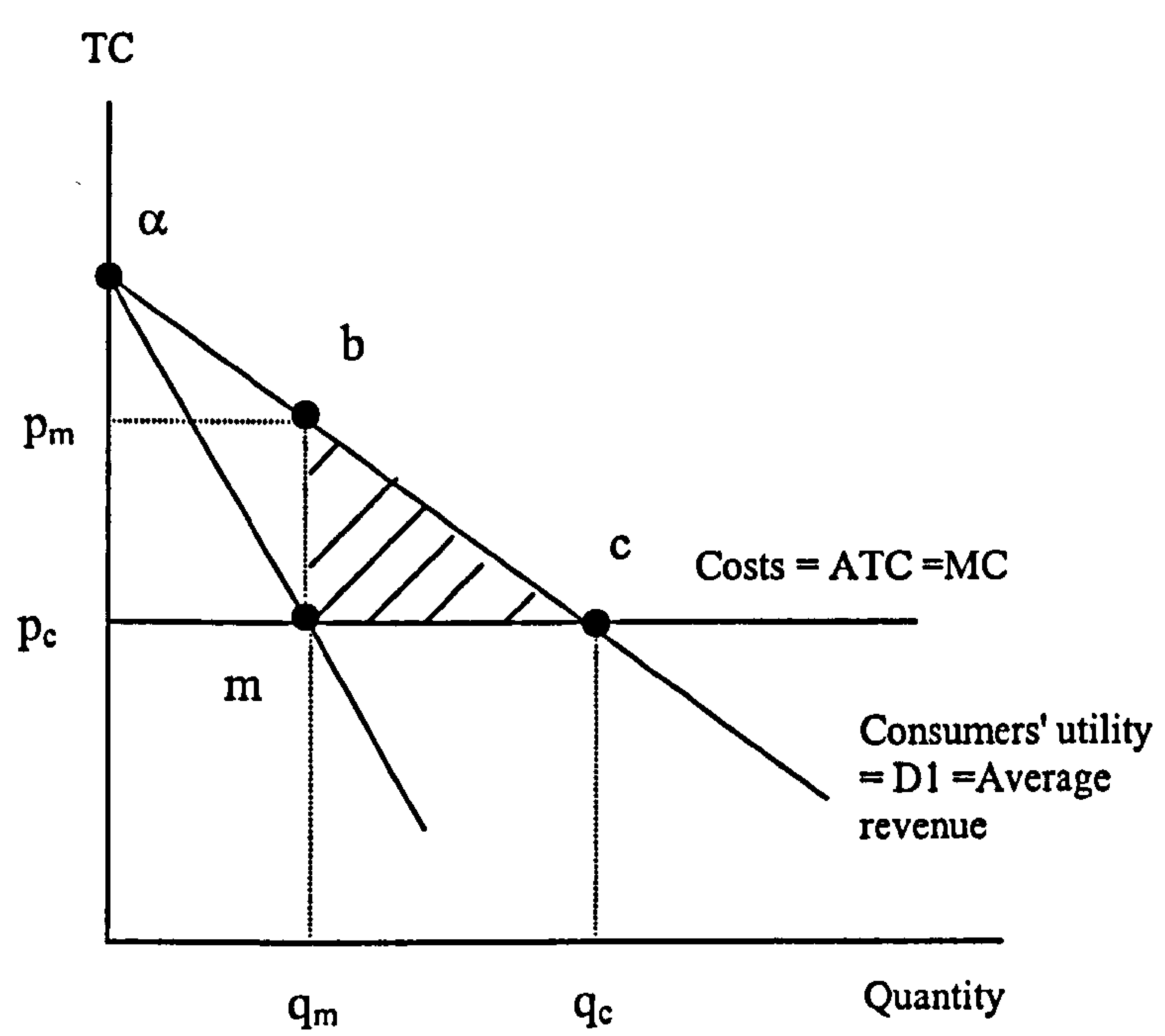

Figure 2-4: Monopoly and Competition compared (Source: Pappa et al. (1991), p.419)

With a constant cost function, point c represents the long-run competitive equilibrium, where price is equal to marginal cost. Point m denotes the equilibrium under monopoly. It is clear from the above figure that under monopolistic conditions consumers will pay higher price ($p_m > p_c$), while the quantity produced will be lower ($q_m < q_c$). This result to the fundamental proposition in economic theory that monopoly is 'bad' compared with perfect competition.

Now, in terms of consumer welfare, the demand curve *(D1)* represents the value that consumers place on the product, and thus its social utility. For simplicity we assume that unit costs are constant and therefore correspond to marginal cost, i.e. $ATC = MC$. Consequently, under competitive market equilibrium price will corresponds to p_c and output to q_c. This gives a consumer surplus corresponding to triangle α -p_c-c.[43] On the other hand, under monopoly market structure consumer surplus would be restricted to α -p_m -b. Moreover, there is also a producer surplus corresponding to the area of monopoly profit, given by p_m -p_c -m -b. If we further assume that the private cost production is a fair measure of social cost, then social welfare increase as long as consumption utility, measured by the demand function, exceeds marginal cost. Under monopoly, equilibrium output is below this level. Hence, if the market harmonises to competitive structure and settles to point c, welfare would be maximised. This, is turn, creates a new consumer surplus corresponding to the triangle b- m - c, which coincides with the net gain for society. If this is the case, area b- m - c is regarded as a dead-weight welfare loss because it is a loss under monopoly structure.

Under imperfect competition, policy makers should therefore enhance market competition. However, private costs of producers and the valuation of consumers, might not measure satisfactorily social costs and benefits, due to economies of scale. By definition, economies of scale suggest that larger firms produce at a lower cost. If we accept that economies of scale might be a significant factor, what are the implications for our analysis? The answer is not clear-cut, but it is possible that the fundamental conclusion that monopoly leads to a higher price and lower output no

[43] The consumer surplus is the sum of the excess consumer valuation expressed by the demand function over the price paid, p_c, on all inputs up to q_c.

longer holds. Hence, despite the prima facie evidence against monopoly, in some cases it is necessary to allow for the potential gains from scale economies.

Furthermore, the existence of fewer firms in the market results in lack of competitive pressure that, in turn, permits organisational slack to develop. Less competition suggests less motivation for managers to control cost at optimal levels. Hence, firms are operating at levels above the point that is economically efficient, exhibiting X-inefficiency. The latter is assumed to be increase with market power [Berger and Hannan (1993,1997, and 1998)]. This generates an important gap between potential economies of scale and realised economies of scale due to of X-inefficiency arising from market power.

Until now, there is no comprehensive theory on the best market structure for an industry. Each industry should be considered separately, and examined in reference to its own organisational behaviour.

The need to test market structure and performance relationship in banking is based on the notion that effectiveness of banks in serving deposits and credit needs of an economy is in some way related to the structure and organisation of banking industry. There are two basic policy issues that evaluate optimality in banking structure in terms of both cost and availability. The first one is efficiency and the other is minimisation of failure. In economic theory, efficiency is related to competition, i.e. in a more competitive environment industries are more efficient, as inefficient firms are forced to leave the market. On the other hand, under monopolistic conditions, inefficient firms can still operate. Accordingly, the objective of a failing proof banking industry is being served under monopolistic condition, whereas a competitive industry provides the conditions for more efficient firms. However, the existence of economies of scale complicates the choice between the two market regimes. Since economies of scale arise when there is an increase in bank size, fewer competitors should exist in the market.

In general, if the performance of banks is not affected by changes in industry structure, regulatory authorities need not to intervene. In the opposite case, regulators must intervene to maintain the well functioning of the industry and the economy.

Moreover, market structure and performance relationship has several implications for consumers and banks. When the market is characterised by high degree of competition, banks work towards increasing efficiency to become more competitive and viable. Consequently, higher efficiency results in lower costs and higher profit. From the consumers point of view higher competition implies better services as well as lower prices. On the other hand, when the market is characterised by monopolistic conditions, banks might engage in some activities that raise cost more than revenue. In turn, consumers will be charged with higher loan rates and lower deposit rates.

2.6 Conclusion

In this chapter, we analysed the theory of the banking firm. The first step was to examine the reasoning for the existence of banks as opposed to other financial intermediaries. Studies on banks emphasise their role in providing liquidity by transforming illiquid assets into liquid liabilities, acting as delegated monitors of investors, as well as their traditional function of providing payments and portfolio services. The social value associated with some of these functions provided by the banks, such as the provision of liquidity and payments services, exceeds their private value. Interruption of these services is costly as the production process is severely affected and assets are liquidated prematurely. Given the central role of banks for the smooth functioning of an economy, extensive failures among banks could seriously obstruct other economic sectors. With the introduction of regulations probabilities of bank failure are minimised.

After examining the rational for regulation, we concentrated on the definitional problem of bank input and output. We reviewed complications arising when one attempts to define and measure bank output and input, with special attention paid to production intermediation approaches. The literature has exposed considerable difficulties in the definition and measurement of bank input and output, in as much as financial institutions are multi-product firms producing services rather than physical products.

Next, key competitive issues related to banking markets were examined. First, we discuss extensively scale and scope economies in banking. Scale economies suggest that larger firms produce output at lower average cost than smaller firms. Economies of scope are present when joint output of a single firm is greater than potential output of two different firms each producing a single product. Next, we discussed efficiency differences (X-inefficiencies). Studies suggested that differences in managerial ability to control costs or maximise revenues seem to be larger than costs effects of scale and scope production.

The final section of this chapter looked at the market structure and performance relationship. The effectiveness of banks in serving deposits and credit needs is related to the structure and organisation of the banking industry. Two basic policy issues evaluate which banking structure best serves the economy, in terms of both cost and availability. The first one is efficiency and the other is the minimisation of failure. Efficiency is related with competition, the more competitive the environment is, the more efficient is the industry. On the other hand under monopolistic conditions, firms that are inefficient can still operate. Therefore, the objective of failing proof banking industry is being served under monopolistic condition, whereas a competitive industry provides the conditions for more efficient firms.

CHAPTER 3: Scale and Scope Economies in UK Banking.

3.1 Introduction.

The primary objective of this chapter is to assess UK banks' ability to realise economies of scale and cost complementarities in joint production by modelling the structure of costs and production for a multi-product institution. The main motivation emerges from the need to assess the effect of all the fundamental changes occurred in the cost structure of the UK banking industry. Given that the major changes, both nationally and internationally, took place during 1980s and 1990s, the current study employ UK data from 1981 to 1995.

Knowledge of the optimal size and product mix is important for decision-making by banks and by regulatory authorities. Given the rapid developments in the British banking during the last two decades, new opportunities and new powers have been granted to banks. The evolving structure of the industry will depend on the extent of size (scale) and product mix (scope) available to them. Taking into account the increasing pattern of competition and concentration within the industry, institutions that adopt the most cost-efficient size and product mix will be in a position to exploit relative cost advantage and will continue to grow. On the other hand, cost inefficient firms will no longer be competitively viable.

The existence of scale economies determines, to a large extent, market structure. An industry that exhibits substantial scale economies is likely to result in an oligopolistic market structure in the long-run [Scherer (1980)]. Furthermore, if scope economies are present, firms may expand their activities by adding new products and services in established markets. Scale and scope economies, favour mergers and acquisition, which in turn lead to a more concentrated financial system. Accordingly, the existence of such economies is of interest to the government for regulation of markets. Information obtained, improve government policy by assessing the effect of regulation, mergers, and the structure of the market. Further, restrictive regulation impending growth and diversification of banks could lead to operational inefficiencies in the form of diseconomies of scale and/or joint production.

Many studies that investigated economies of scale and scope in banking markets over the last thirty years. Nevertheless, most of the empirical literature on scale and scope economies investigates the cost structure of the US market. Results differ depending on the size of banks used in the data sample, but in general, studies tend to conclude that large bank over-performed small ones in terms of scale efficiency. For the European market little work has been done, and although existing evidence suggests the presence of scale economies, there is no agreement to the level of output at which economies are exhausted. Some studies examine the existence of scope economies with varying conclusions: overall, findings suggest stronger scale economies compared to scope economies. Therefore, it is clear that more research should be undertaken, for the cost structure of the European banking industry.

Section 2 refers to the alternative methodologies utilised for the estimation of production and cost definition. Section 3 presents the different measures of scale and scope economies employed in the banking literature. Section 4 provides the possible sources for the existence of scale and scope economies. Section 5 gives a detail description of the literature review. Section 6 provides a detail analysis of the methodology applied. Section 7 discusses the data set utilised. Section 8 presents the input and output specification. Section 9 gives a descriptive analysis of the empirical results, and section 10 summarises the study and presents conclusions.

3.2 Production and Cost Functions: Alternative Methodologies

The relationship between bank's input and output can be examined via a production function or its appropriate cost function. The relevant literature has employed different functional forms so as to capture the multi-product nature of banking firm. This section provides an overview of the different methodologies applied in the banking literature.

In general, the problem of the firm is to maximise profits subject to a given technology. In the neo-classical formulation, technology is summarised by a production function, indicating the maximum output attainable for alternative combinations of production factors.

The production function of a firm, producing a single output from two inputs can be represented as:

$$y = f(x_1, x_2) \tag{3.1}$$

where:

y = denotes maximum possible output,

x_1, x_2 = denotes inputs levels.

In most empirical applications the production function gives output y as a function of capital (K), and labour (L). Hence:

$$y = f(K, L) \tag{3.2}$$

Equation (3.2) may exhibit returns to scale (decreasing, constant, or increasing) at particular points. Returns to scale indicate the response of output to an equiproportionate change in both inputs. In order to have returns to scale phenomena, Equation (3.2) must be homogeneous of degree n. Then depending on whether n, is less than, equal to, or greater that unity, equiproportionate increases in input will lead to a less than proportionate, equiproportionate, or greater that proportionate increase in output. Consequently, there are decreasing, constant, or increasing returns to scales respectively. Going back to Equation (3.2), returns to scale can be measured by:

$$f(\lambda K, \lambda L) = \lambda^n (K, L) \tag{3.3}$$

where:

λ = denotes equiproportionate change in factor inputs,

n = denotes the degree of homogeneity.

The most widely used production functions for empirical estimations is the **Cobb-Douglas** production function which take the form:

$$y = AK^{\alpha}L^{\beta} \tag{3.4}$$

where A, α, and, β are fixed parameters. The Cobb-Douglas function is homogeneous of degree $\alpha + \beta$, since:

$$y(\lambda K, \lambda L) = A(\lambda K)^{\alpha}(\lambda L)^{\beta} = \lambda^{\alpha+\beta}y(K,L) \tag{3.5}$$

Therefore, if $\alpha + \beta > 1$, there are increasing returns to scale, if $\alpha + \beta = 1$, there are constant returns to scale, and if $\alpha + \beta < 1$, there are decreasing returns to scale.

The Cobb-Douglas is linear in the logarithms of the variables. Considering cross-section studies, the Cobb-Douglas function for the i^{th} firm, after taking logarithms and adding a stochastic disturbance term ε_i to account for variations in the technical or production capabilities of the firm i becomes:

$$\ln y_i = \ln A + \alpha \ln K_i + \beta \ln L_i + \varepsilon_i \tag{3.6}$$

where A includes the effect of technology. As regards parameters α and β it is assumed that they are the same for all firms, and any differences among firms are captured by the disturbance ε_i, which is assumed to have zero mean, and σ^2 variance. Finally, parameters α and β measure the elasticities of output with respect to capital and labour, taking values between zero to unity.

However, the Cobb-Douglas specification allows only for uniform scale characteristics. Thus, the cost function is restricted in reflecting, either increasing, constant, or decreasing returns to scales. No combinations of the scale are allowed, as required by e.g. a U-shaped cost curve. A further restriction imposed by the Cobb-

Douglas specification is that the elasticity of factor substitution[44] equals unity, which limits substitution between production factor.

Another major problem in the use of Cobb-Douglas specification is the disregard for the existence of a dual relationship between the production function and cost function. The cost function is dual to the production function in that it provides an alternative, yet equivalent description of the firm's problem. A firm maximises its profits either by maximising the amount of output produced, given its resources, or equivalently, by minimising its cost. Maximisation of output, or minimisation of cost, are two alternative approaches to profit maximisation. In order to properly estimate cost, it is necessary that the cost function reflects all properties of the dual production function.

Considering now the cost function, as in Nerlove (1963), total cost can be given:

$$TC = p_1K + p_2L \tag{3.7}$$

where p_1 and p_2 are capital and labour prices, respectively. Minimising total cost subject to the production function in Equation (3.4) gives :

$$p_1K/\alpha = p_2L/\beta \tag{3.8}$$

As Nerlove noted, there are problems with the estimation of relationships between input and outputs arises whenever prices paid for factors vary from firm to firm, because independence among production efficiency, output level, and factor prices no longer holds. Therefore, it is preferable to derive estimates of structural parameters from estimates of reduced-form parameters.

However, Nerlove did not estimate each reduced-form equation separately, but (since all data concerned total cost) he estimated the total cost function in the form of a

---

[44]The elasticity of factor substitution is defined as $\sigma = \dfrac{d(K/L)}{K/L} / \dfrac{d(r/w)}{r/w}$, where r, and w are prices of capital and labour respectively. It measures the proportionate rate of change of the input ratio divided by the proportionate of change in the ratio of factor marginal products in the Cobb-Douglas production function.

linear combination of reduced form equations. Thus, by substituting reduced form expressions for p_1 and p_2 into equation (3.4), he obtained:

$$TC = \gamma\, y^{1/r} p1_K^{\alpha/r} p2_L^{\beta/r} \varepsilon^{-1/r} \tag{3.9}$$

where,

$r = \alpha + \beta$,

$\gamma = r(A\alpha^{\alpha}\beta^{\beta})^{-1/r}$.

Parameter γ is the constant and the r is the returns-to-scale parameter, equal to the sum of output elasticities with respect to capital and labour. Because cost function (3.9) is directly derived from minimising the cost relationship (3.7) subject to the production function (3.7), duality between cost function and production function is assured.

The cost function can be expressed in log-linear form as:

$$\ln TC = \ln\gamma + \left(\tfrac{1}{r}\right)\ln y + \left(\tfrac{\alpha}{r}\right)\ln p_1 + \left(\tfrac{\beta}{r}\right)\ln p_2 - \left(\tfrac{1}{r}\right)\ln u \tag{3.10}$$

Further, since $\alpha + \beta = r$, coefficients on $\ln p_1$ and $\ln p_2$ must add to unity; that is $\frac{\alpha}{r} + \frac{\beta}{r} = 1$. Then equation (3.10) can be rewritten as:

$$\ln TC - \ln p_1 = \ln\gamma + \left(\tfrac{1}{r}\right)\ln y + \left(\tfrac{\beta}{r}\right)(\ln p_2 - \ln p_1) - \left(\tfrac{1}{r}\right)\ln\varepsilon \tag{3.11}$$

Equation (3.11) can be estimated to give a measure of scale economies. However, a problem may arise with uniform scale economies, as U-shape cost curve is not allowed, and the elasticity of substitution equals unity.

A more general specification is the Constant Elasticity of Substitution (CES) production function [Arrow *et al.* (1961)], which allows for any degree of substitutability between two input factors. The typical CES has the form:

$$y = A[bK^{-\theta} + (1-b)L^{-\theta}]^{-z/\theta} \tag{3.12}$$

where z denotes the returns to scale, and θ is the substitution parameter. Elasticity between K and L is constant and is measured by $\sigma = 1/(1 + \theta)$.

Still, there is an obvious need for the development of variable elasticity of substitution (VES) specification. VES functions were developed in the early 1970s, and the major problem was to find a functional form that allowed a variable elasticity of substitution and that could be easily estimated.

A general form of VES functions was presented by Christensen *et al.* (1973), known as 'transcendental logarithmic' or translog production and cost functions.[45] Translog models allow the estimation of production or cost function with more than one output and more than two inputs. Moreover, they are flexible, in that they allow any degree of substitutability at all levels of employment. Finally, they allow estimation of a U-shaped cost curve.

The general form of the translog production function can be expressed as:

$$\ln y = a_0 + \sum_{i=1}^{n} a_i \ln x_i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \ln x_i x_j \tag{3.13}$$

where $a_{ij} = aji$, and i, j and x_i are n quantities of inputs.

The big advantage of (3.13) is easy estimation. In general, this function is quiet flexible in approximating arbitrary production technologies in terms of substitution possibilities. It provides a local approximation to any production function, and has also the property that implied returns to scale implied are not the same across all values of input and output.

[45] Christensen *et al.* (1973) developed a single output technology translog function. The multiple case was presented by Burgess (1974) and Diewert (1974).

Following the duality principle, that cost functions can be obtained directly from production functions [for proofsof duality principles, see Diewert (1974, 1982), Fuss and McFadden (1980), Nadiri (1982)] the following translog cost function can be derived:

$$\ln TC = a_0 + \sum_{i=1}^{m} a_i \ln y_i + \sum_{i=1}^{n} \beta_i \ln w_i + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \delta_{ij} \ln y_i \ln y_j + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} \ln w_i \ln w_j + \sum_{i=1}^{n}\sum_{j=1}^{m} \rho_{ij} \ln w_i \ln y_j$$

(3.14)

where w_i denotes price of inputs. There are $m + 1$ independent α_i parameters, n independent β_i, $m\ (m + 1) / 2\ \delta_{ij}$ parameters, since it is assumed that $\delta_{ij} = \delta_{ji}$ for $1 < i, j < m$, $n\ (n + 1) / 2\ \gamma_{ij}$ parameters, since it is assumed that $\gamma_{ij} = \gamma_{ji}$ for $1 < i, j < n$, and mn independent p_{ij} parameters in the translog cost function defined by equation (3.14).

The translog cost function can be regarded as a second-order approximation to any arbitrary cost function, is linearly homogeneous in input prices, and like the translog production function permits the nature of the return-to-scale implied to vary with the levels of inputs and outputs.

In general, the translog cost function is preferred in studies of bank performance because it allows, first, the estimation of U-shaped average cost curve, and, second, the derivation of scale economies. Further, it does not impose theoretical restrictions and permits economies to vary by size of banks. Most important, it allows the estimation of effects of multi-product operations on costs and subadditivity[46] on unit and marginal costs.

However, when one output becomes zero the translog specification does not allow the evaluation of scope economies and product specific economies of scale,. To overcome this problem, Caves *et al.* (1980) developed a hybrid translog cost function,

[46] A cost function is subadditive when single-firm production of the output mix $Y_i = (Y_1, Y_2, \ldots, Y_n)$ is less costly than production of the output mix of Y_i using any combination or division of Y_i among group of firms, holding technology and input prices constant.

which is a generalisation of the translog function. The difference between the translog and the hybrid function is that in the latter the log of output is replaced by a Box-Cox transformation, that is:

$$y^{*} = \frac{(y^{\lambda} - 1)}{\lambda} \qquad \lambda \neq 0$$

$$y^{*} = \ln y_i \qquad \lambda = 0 \tag{3.15}$$

When λ approaches zero, the hybrid cost function approaches the translog function. When λ equals one the cost function becomes semilog. Therefore, by treating λ as an additional unknown parameter the model becomes more flexible.

3.3 Economies of scale and scope in a multi-product firm

Scale economies in the single product firm exist if total cost increases less than output. Consider a firm with the cost function, $TC = f(y)$, where y is output. Then the average total cost can be derive as $ATC = f(y) / y$, and the marginal cost is $\partial TC / \partial y$. Consequently, economies of scale can be defined as:

$$SE = \frac{ATC}{MC} = \frac{\partial y}{Y(\partial TC / \partial y)} \tag{3.16}$$

which is the elasticity of cost with respect to output. Hence, when SE $\geq$ 1, SE = 1, and SE $\leq$ 1, we have increasing, constant or decreasing returns to scale, with the derivative of average cost with respect to output being negative, zero, or positive, respectively.

The multi-product nature of banking institution renders the analysis and interpretation of economies of scale and scope more complex since average cost is defined for a single product process. Conventional average cost curves cannot be derived for multi-product firms unless all products are aggregated into a single index, which in most cases is not feasible. In the banking literature the behaviour of costs is examined in proportion to bank's output. If each product in the bundle increases by a given percentage, the quantity of the composite commodity increases by the same percentage. This approach is known as the ray average cost (RAC) [see Baumol (1977), and Baumol, *et al.* (1988)]. *Ray average costs* are a natural generalisation of single product average cost and are defined as follows:

$$RAC(y) = \frac{TC(y)}{\sum y_i} = \frac{TC(tq^0)}{n} \tag{3.17}$$

where q^0 is the unit bundle for a particular mixture of outputs, and n is the number of units in the bundle. Accordingly, $y = nq^0$. The degree of scale economies at q^0 is defined as the elasticity of output with respect to cost, equivalent to $1/(1 - \varepsilon)$, where it is greater than, less than, or equal to one as returns-to-scale are locally increasing, decreasing, or constant, and as the slope of the RAC curve is negative, positive or zero, respectively.

A number of researchers suggested that the appropriate measure of *overall economies of scale* in a multi-product case is the sum of individual output cost elasticities with respect to output, holding the product mix constant. That is, total change in cost equals the sum of differential changes in the level of *m* outputs:

$$OSE = \sum_{i=1}^{m} \frac{\partial \ln TC}{\partial \ln y_i} \qquad \text{for m} = 1,\ldots,\text{i} \tag{3.18}$$

If $OSE < 1$, overall marginal costs are decreasing and, therefore, production is subject to increasing returns to scale, and thus economies of scale. If $OSE > 1$ there are decreasing returns to scale (implying diseconomies of scale) and, finally, if $OSE = 1$ there are constant returns to scale.

Overall economies of scale explain the behaviour of cost as output increases or decreases for a given bundle of outputs. However, the behaviour of total cost in not explained when there is a change in the mixture of output. Changes in total cost as output of one-commodity changes are measured with *product specific economies of* *scale* (PSES) [Panzar and Willing (1988a), Baumol *et al.* (1988)]. PSES are evaluated with the use of average incremental cost (AIC). AIC is defined as the extra cost of adding the production of that given product at a specific level of output, rather than not being produced at all, divided by the output of that product. Measurement of PSES for product i at output vector y is given by the ratio of the AIC of the product to its marginal cost (TC/y_i), that is:

$$PSES_i = \frac{AIC_i}{(\partial TC / \partial y_i)} = \frac{IC_i}{\varepsilon_{TC_{Yi}} TC} \tag{3.19}$$

where $\varepsilon_{TC_{Yi}}$ is cost elasticity if i^{th} output, and IC_i is the incremental cost of product i^{th}, given by

$$IC_i = TC(y_1, y_2) - TC(0, y_2) \tag{3.20}$$

where $PSES > 1$, $PSES = 1$, and $PSES < 1$ shows decreasing, constant and increasing returns-to-scale, with respect to output i.

In terms of product mix, the existence of cost savings can be examined via scope economies. Economies of scope between two outputs exist when a twice differentiable multi-product cost function exhibits cost complementarities between the products. Cost complementarities (COMP) exist if the marginal cost of producing one output decreases when the production of the other output increases, i.e:

$$COMP_{ij} = \frac{\partial^2 TC}{\partial Y_i \partial y_j} \tag{3.21}$$

If $COMP < 0$ then cost complementarities are present implying economies of scope, while if $COMP > 0$ we have diseconomies of scope; a value equal to zero suggests cost independence.

There are also two other measures of scale and scope based on the cost subadditivity concept, the expansion path scale economy (EPSCE) and the expansion path subadditivity (EPSUB) [Berger *et al.* (1987)]. Cost subadditivity measures the relative efficiency of large and small firms and takes into account both scale and scope economies simultaneously. *EPSCE* denotes the cost advantage or disadvantage of a large firm, *A*, compared to a smaller one, *B*, by measuring the elasticity of cost with respect to changes in output. This is calculated as:

$$EPSCE = \sum_{i=1}^{n} \frac{(y_n^A - y_n^B)}{(TC(y^A) - TC(y^B))/TC(y^A)} * \partial \ln TC(y^A)/\partial \ln y_n^A \tag{3.22}$$

where $y_n^A, TC(y^A)$ denotes the output and the total cost for firm *A*,

$y_n^B, TC(y^B)$ denotes the output and the total cost for firm *B*,

$y_n^C, TC(y^C)$ denotes the output and the total cost for firm C.

In the second measure, *EPSUB*, a large bank opposed to a smaller one are considered, together with a third, hypothetical bank, *C*, whose output is the difference of the other two banks' outputs. This can be represented algebraic form:

$$EPSUB = \frac{TC(y^B) + TC(y^C) - TC(y^A)}{TC(y^A)} \tag{3.23}$$

The measure gives the relative cost increase or decrease arising from the production of large bank's output by the small bank and the complementary bank. When EPSCE<1, there exist economies of scale, whilst EPSCE>1 indicates diseconomies. When EPSUB<0 then firm B is not competitively viable. Consequently, it may be driven out of a competitive market by the combination of firms at A and C.

3.4 Sources of economies of scale and scope

As the concepts of economies of scale and scope have been analysed in chapter 2, we discuss here the possible sources for the existence of such economies. Literature concentrates on four major causes for scale economies. The first one suggests that economies of scale are the effect of a more efficient use of some or all inputs with an increase in the volume of output [Benston (1965a), Bell and Murphy (1968), Scherer (1980)]. Firms may have unavoidable excess capacity of some inputs, so that an increase in output may not require a proportionate increase in input. Specifically, some inputs might be wholly or partly indivisible by output. The existence of indivisibility may help reduce costs per unit of output as the output level increases.

Second, increased size may allow a more efficient organisation of resources. A relatively large bank allows for greater input and process specialisation, e.g. in a large bank, one employee is assign to one task, whereas in a smaller bank one employee is often assign to a variety of tasks.

Third, large banks have the economic advantage, over smaller ones, for some types of technical innovations such information technologies. Hence, according to asset size banks could employ different compositions of inputs with varying degrees of efficiency.

Fourth and last, Kolari and Zardkoohi (1987) state that the law of large number of accounts for certain economies of scales. Large banks may be not required to hold as much cash balances as small banks (in proportion to their anticipated transactions). Since holding cash balances is costly, larger banks incur lower cost than smaller banks, to the extent that the law of large numbers smoothes transaction demand[47].

Possible sources of scope economies in banking, as suggested by the literature, come forth from banks' multi-product structure [Baumol *et al.* (1988, p 75-79), Berger *et al.* (1987, p. 503-504), Mester (1987a, p. 17-18). First, information economies exist with the join use of information for a certain customer by different divisions. Second, fixed costs could be spread over an expanded product mix. Third, asset diversification across different assets group may reduce portfolio and interest rate risks. Fourth, there is a reduction in customers' cost by doing business in a wide range of product and services at the same branch.

[47] For a detail analysis see Kolari and Zardkoohi (1987)

3.5 Literature Review

3.5.1 Early empirical studies

Early cost studies utilised financial ratio analysis to examine the existence of economies of scale in US banking industry, by use of data gathered from balance sheet and income statements. These studies can be distinguished into two categories based on output specification. The first category includes studies that measure bank output in terms of earning assets, while the second category comprises of studies that uses total assets as a measure of output.

3.5.1.1 Earning Assets as a Output Measure

The first studies that documented cost differences between bank of different size were Alhadeff (1954), and Horvitz (1963). Both studies defined output as the ratio of loans plus investments to total assets (earning assets), and cost included operating and interest costs. With the utilisation of a simple cost function, scale economies occurred if average costs decline as average cost increases. They find evidence of constant returns to scale for mid-sized banks, while small and large size banks exhibited increasing returns to scale. Furthermore, they reported that branch banks had higher average costs than unit banks.

3.5.1.2 Total Assets as Output Measure

The major criticism of the aforementioned studies was that output measure included only earning assets, whereas other assets, such as trust operations, were excluded. This omission led to an increase of larger banks' average unit cost. To overcome this limitation, Schweiger and McGee (1961) and Gramley (1962) used total assets as a measure of bank output. Both studies employed multiple regression analysis and find evidence that average costs were decreasing as the size of bank increased, and therefore larger banks are found to have a cost advantage over small banks.

From the above it is clear that earlier cost studies for the US banking industry produced conflicting results. Studies that used earning assets as a measure of output find that medium sized banks faced constant unit cost, and small and large banks faced a lower average cost. On the other hand, studies that employed total assets as measure of output found a declining average cost for all ranges of output size.

3.5.2 Cobb-Douglas cost function

Subsequent studies proceed one step further, and, instead of simple statistical models based on ratio analysis, model bank cost with the use of the Cobb-Douglas cost function. A further advance of these studies is the attempt to model the multi-product nature of banks with the inclusion of multiple outputs. Table 3-1 summarises the results of a number of empirical studies in banking for the US markets using Cobb-Douglas cost function. The general conclusion that emerges from the early research in American banking industry was the existence of economies of scale for most individual services offered by banks, varying according to the banks internal organisation. At that time this conclusion was crucial, as it implied a monopolistic structure of the industry. However, at the end of 1970s this conclusion began to be questioned on both theoretical and methodological grounds, and by the 1980s was seriously held in doubt.

The first criticisms came from Baumol (1977) who indicated that economies of scale where neither a necessary nor a sufficient condition for the existence of monopoly in a multi-product industry such as banking. The importance of economies of scale in one-output case is that the existence is required to have a natural monopoly. Nevertheless, in the multi-product case, scale economies only describe the technical gains from increases in scale of operation, rather than gains originating from the production of several goods jointly. Consequently, it became necessary to introduce new indicators in the analysis of the banking costs. The main purpose of these indicators was to determine whether monopoly was a natural result of unregulated markets, since deregulation had already started. Because scale returns were inadequate as indicator, economies of scope or joint production were further examined with the origin of the latter being the combined use of the same input to produce different output or services. Also, earlier studies employed the Cobb-Douglas methodology, which does not allow a U-shape functional form, no does it allow testing for the existence of economies of scope. Hence there was a need for a more flexible functional form.

Table 3-1: Cobb-Douglas cost function studies in the US banking industry.

Authors	Data	Output and Control Variables	Conclusions
Benston (1965a)	Gross-section data from the FCA for 80 banks. 1959-1961	Number of demand deposits, time deposits, mortgage loans, instalment loans, business loans, and securities. CV: Average account sizes, branching and mergers dummies, price and total assets.	Economies of scale: demand deposits and mortgage loans. Diseconomies of scale: time deposits, instalment loans. Banks size does not it self provide a cost advantage.
Benston (1965b)	Same data set as in Benston's (1965a).	Same variables as in Benston's (1965a).	Banks with three or fewer branches, experience cost benefits, and as the number of branch offices increases, demand deposit, instalment loan, and occupancy cost increase as well. One and two branch banks have cost characteristics consistent to those of unit banks.
Bell and Murphy (1968)	Gross-section data form the FCA consisted of 283 banks. 1963-1965	Same output variables as in Benston's (1965a).	Economies of scale for demand deposits and real estate loans, and slight diseconomies of scale for time deposits and instalments loans. Branch banks were more cost efficient than unit banks, while the cost functions' of the three types of banks, were very different.
Murphy (1972)	Gross-section data from the FCA consisted of 967 banks, for 1968.	Same output variables as in Benston's (1965a).	Banks are characterised with constant returns to scale.
Schweitzer (1972)	Call Reports and Condition Data on the 9th District	Lending output index: measured as total loan revenues plus revenue from	There is a U-shaped cost curve. Scale economies are exhausted for the most part after \$3.5 million in assets,

	banks for 1964	other sources. Dummy variables: location, bank holding company, Federal Reserve membership.	and diseconomies exist beyond $25 million in assets. Constant returns to scale in the $3 to$25 million asset range.
Daniel, Longbrake, and Murphy (1973)	Gross-section data from the FCA consisted of 967 banks, for 1968.	Number of demand Deposits CV: Average size variable, annual wage rate, rental rate, number of branches.	Only large banks can improve operating efficiency with the use of technology. Banks with fewer than $18 millions demand deposit accounts should use conventional accounting systems
Kalish and Gilbert (1973)	Data from FCA consisted of 898 banks for 1968.	Lending output index: measured as total loan revenues plus revenue from other sources.	Banks affiliated with a holding company have greater costs efficiency than branch banks at lower output levels, but he reverse is true at higher output levels. Further, cost curves are U-shaped
Longbrake, and Haslem (1975)	Gross-section data form the FCA consisted of 967 banks, for 1968.	The same approach as Daniel Longbrake, and Murphey.	Differences in cost function shapes between banks with various legal forms of banks' organisation; The number of branch bank offices had little effect on the cost of producing demand deposit services.
Mullineaux (1975)	1970 data from Federal Reserve Bank of Boston, New York and Philadelphia.	The same approach as Benston (1965) and Bell and Murphy (1968) CV: dummy variables for branching	Evidence for the existence of economies of scale for unit banks; for branch banks economies of scale were small.

3.5.3 Translog Cost Function

Earlier studies in banking faced three important limitations. First, they focused on individual bank functions; hence, total cost of banking operations was not addressed. Second, an average cost curve that could take a U-shape over the full range of banks was not fitted, either because larger banks were omitted or because of the use of Cobb-Douglas function. As a result, the optimum (minimum) cost size of a bank, or office, could not be determined. Therefore, they produced inappropriate efficiency comparisons between unit and branch banking [Benston *et al.* (1982)]. Last but not least, no tests for the existence of economies of scope were undertaken. In an attempt to overcome these problems, studies in the 1980s used the translog methodology in examining scale and scope economies in the US banking industry. As mentioned earlier, the translog cost function allows the estimation of U-shaped average cost curve and, also, it does not impose theoretical restrictions. Indeed, results from translog studies suggested that average cost curves in the US banking market are U-shaped.

Table 3-2 summarises the results of a number of empirical studies in banking for the US market using the translog methodology. The major conclusions from these studies are that bank's average operating cost have U-shape form, hence the Cobb-Douglas functional form should not be applied to cost studies. Moreover, although earlier studies find that economies of scale are exhausted at low levels of output, later studies reported mixed results. For instance, Benston *et al.* (1983) concluded that economies of scale exist for all sizes of branch offices, except branches with more than $400 millions of deposits. The same authors find that in the case of unit banks, economies of scale exits only for the first $50 millions of deposits. Another major contribution, from those studied in the literature, was the test for the existence of economies of scope. Some evidence of cost complementarity is found, but an important drawback of these studies is the utilisation of FCA data, which ignores large banks and only concentrates on smaller institutions [Hunter *et al.* (1990)]. Hence, the results from these studies are not useful for estimating the effects on operating costs of banks with total deposits over of more than $1 billion.

Table 3-2: Translog cost function studies in the US banking industry.

Authors	Data	Conclusions
Benston, Hanweck, Humphrey (1982b)	Gross-section data from the FCA 1975-1978	Average operating cost curve are U-shaped or upward sloping for both unit banks and branch banks. Unit banks above $50 million of deposits exhibit diseconomies of scale, branch banks exhibit small economies of scale. Optimum size of a bank is, between $10 to $25 million of deposits.
Benston, Berger, Hanweck, Humphrey (1983)	FCA data 1978	Scale economies exist for all branch offices. At firm level no scale economies or diseconomies exits. For unit banking economies of scale exists for less than $50 millions of deposits, diseconomies of scales at $200 millions of deposits. Evidence of pairwaise cost complementarities.
Clark (1984)	Data from Call Reports of Income and Conditions for 1205 unit banks, from 1972 to 1977	Results are insensitive to the choice of output. Economies of scale in banking appear to be small; exits only for small-sized banks. An increase in size is unlikely to improve the operating efficiency of a bank.
Gillian and Smirlock (1984)	Data from Call Reports of Income and Conditions by the FRB of Kansas for more 2,700, form 1973 to 1979	Scales economies for banks with less than $10 millions of deposits, diseconomies of scale for banks beyond $50 millions of deposits; bank costs are characterised by a U-shaped cost function. Cost of producing one output depends on the level of other outputs.
Lawrence and Shay (1986)	Gross-section data from FCA, from 1979 to 1982	Scale economies at all size levels for branch banks; significant scale economies only for smallest unit banks. Significant economies of scope between deposit and loans, deposit and investments, diseconomies of scope between investment and loans.
Kolari and Zardkoohi (1987)	Gross-section data from FCA, from 1979 to 1983	Economies of scale for banks with up to $50 million of deposits, diseconomies for banks beyond the level of $50-100 million of deposits, except 1983. Evidence for scope economies. Flat cost curves for unit banks, U-shaped or upward-sloping for branch banks

Berger, Hanweck, and Humphrey (1987)	FCA data for 413 branching state banks and 214 unit state banks,1983	Slight scale economies at branch level and slight diseconomies of scale at firm level. Unit banks exhibit large diseconomies of sale for large banks. Evidence of diseconomies of scope
Gropper (1991)	Gross-section data from FCA, from 1979 to1986	Economies of scale exist beyond small levels of output for years prior to 1982; significant economies of scale in post-1982 years for branch and unit banks. Degree of scale economies increases over the period.

3.5.4 Recent evidence

The internalisation of banking services and the deregulation pattern raise new policy issues. Hence, recent studies examine the effect of these changes on the cost of banking industry and the viability of banking institutions on the new and more concentrated and competitive markets. In this section we report evidence from studies that examined the effect of deregulation on the cost structure of the banking markets, evidence from European and international studies, as well as evidence from studies that applied alternative methodologies.

3.5.4.1 Deregulation: their effect on the production and cost structure of banking industry

During the 1990s a number of studies concentrated on the debate on the optimal structure of the banking system and banking regulation. Regulatory changes, which in turn led to changes in banks' balance sheet structure, aimed to make banks more open to competition, and to cope with new kind of risk exposures.

With deregulation, bank powers have expanded allowing them to offer a number of non-traditional activities, and compete directly with other financial institutions in product markets. Nowadays, banks engage in a wide range of non-traditional activities such as loan sales and purchases, securitisation of assets, and off-balance sheet commitments (e.g. derivatives). Traditional bank activities involved a firm link between the asset and liability of the bank's balance sheet. Banks were taking in deposits and lending them out as loans to be held until maturity. However, the emergence of non-traditional activities has led to unbundling of the asset and liability sides of the bank. This, in turn, led to changes in production and cost structure of the banking industry [Mester (1992)].

The new structure that characterises banking markets nowadays generates an important question: do banks improve efficiency due to product mix and/or due to larger scale of operation? Empirical studies produce mixed results. Mester (1992) compares the cost of traditional banking activities of originating and monitoring loans with costs of non-traditional activities of loan selling and buying, and finds that banks operate at a scale slightly than minimum efficient. Also, results suggest that there are significant diseconomies of scope between traditional and non-traditional activities. This, in turn, suggests that banks could benefit from specialisation by separating most of their traditional activities into a different bank. Jagtiani *et al.* (1995) include off-balance-sheet (OBS) products as output in the cost function to measure their effects on scale economies and cost complementarities. They find that OBS activities have none or small impact on scale estimation; results, in general, suggest no scale economies. In addition, no evidence of cost complementarity is found for most combinations of output, except for the case of OBS foreign currency products and guarantees, or aggregate OBS and earning assets. The authors state that the volume of OBS has little or no effect on bank costs.

An additional study that examines the effects of OBS products on banks cost efficiency is Jagtiani and Khanthavit (1996). This study examines, apart from OBS, the impact of risk based capital requirements on bank costs. In general, results indicate that the introduction of risk-based capital requirements, from their announcement to the actual implementation, expose significant structural changes in banking industry in terms of efficient size and optimal product mixes. In the case of small banks, there is negligible effect in terms of efficient size, whereas for the optimal product mix the effect is greater. Before the implementation of regulation the cost function was in most cases subadditive. After the imposition of capital requirements, product-mix economies disappear since the EPSUB index becomes insignificant. In the case of large banks, after the introduction of regulation diseconomies of scale and scope are found, indicating that these banks are too large to be efficient.

In an attempt to capture the quality of banks' assets and banks' individual risk in the light of the new structure of the industry, Mester (1996) includes two additional variables in the cost function apart from input prices and outputs. The first one is the average volume of non-performing loans, which serves as a proxy for the quality of banks' output, and the other is financial capital, which captures bank's risk. From his results it is clear that banks in the sample are operating at constant returns to scale. In addition, there is no evidence for scope economies.

All the above studies tend to focus on cost. Rogers (1998a) employees revenues and profits as additional measures of bank performance. Contrary to other studies, his results suggest definite advantages to joint production of loans, deposits, credit enhancements, and other non-traditional bank outputs, in terms of both costs and profits. In general, studies that find insignificant results in joint production suggest that banks are already producing an optimal mix of products, and hence they cannot increase efficiency [Jagtiani and Khanthavit (1996)]. On the other hand, in terms of profit efficiency, joint production of traditional and non-traditional activities results in higher profit compare to the case where each output is produced by a separate firm [Rogers (1998a)].

3.5.4.2 European and international Studies

The smaller volume of cost studies on European banking suggest larger scale economies compared with US; however, there is no consensus as to the level of output at which these economies are exhausted. Nevertheless, US and European banking are very different in structure. European banks are more universal, whereas US banks are more specialised as universal banking is restricted by US regulations, whereas in some states branch banking is prohibited (see chapter six for information of the US banking regulations). This might constitute a reason for larger scale economies experienced by the European banks, compared to the US banks. The concept of scope economies has been investigated by few studies, which suggest some scope economies for larger banks in some markets, i.e. Germany, and for smaller banks in other markets, i.e. France.

More specifically, early studies on the French banking system concentrate on scale economies, utilising a standard translog methodology and suggested that overall scale economies are rather limited [Dietsch (1988)]. Martin and Sassenou (1992) and Dietsch (1993) find evidence of both scale and scope economies. The results of the first study suggested that small banks benefit from larger scale and scope economies, while bigger banks incur relatively large diseconomies of scale depending on their output mix and their degree of specialisation. As far as the latter study is concerned, the results indicate economies of scales for all bank sizes, whereas economies of scope are not observed at a high level of output.

Studies on the Italian banking industry suggest economies of scale at all size ranges. These economies tend to increase, as bank size becomes larger [Cossutta *et al.* (1988), Baldini and Landi (1990), Conigliani *et al.* (1991)]. Contrary, at firm level scale economies become small, as they tend to decrease with the increase in banks' asset size. In terms of scope economies, studies do not find any evidence in favour of their existence.

Cost economies studies in the UK concentrate on the building society sector. The evidence indicates economies of scale, although there is no consensus as to what bank size these economies are exhausted. On the other hand, as regards scope economies, studies find no evidence for their existence. Cooper (1980) finds that scale economies for asset size less than £100 million, where as larger societies are subject to diseconomies. Hardwick's (1989) findings suggest that societies with assets under £280 million benefit from scale economies, where those with assets above £1500 million are subject to diseconomies. Drake (1992, 1995) finds mild scale economies for societies in the £120-500 millions size range. On the other hand, no evidence for scope economies is found. McKillop and Glass (1994) find evidence of scale economies for both national and local societies, but only constant returns for those that are regionally based. They also find no evidence of scope economies or of cost complementarities between the provision of mortgage and non-mortgage outputs.

Lang and Welzel (1994, 1996) employ the translog cost function approach and estimate cost economies for German co-operative banks. Their results suggest moderate economies of scale for all size classes, and evidence for scope economies, especially for larger banks. Fields *et al.* (1993) examine the existence of scale economies in Turkish banks for the years 1986 and 1987 with the utilisation of a standard translog methodology. Their results indicate no significant evidence of economies of size for either year, and conclude that their results are similar to other studies for different time periods, data, and types of countries.

Glass and McKillop (1992) investigate the existence of scale and scope economies for 1972 to 1990 in the Bank of Ireland, and find that for the sub-period 1975-1978 the bank is characterised by overall diseconomies of scale. For the whole period product-specific scale economies are found to be decreasing for investments and increasing for loans. Moreover, the results suggest diseconomies of scope.

All previously mentioned studies are based on a single country; yet, commercial banking activities experience globalisation for the past years. An international study performed by Allen and Rai (1993) estimate a global cost function with a data set of 150 banks from 15 counties. Results indicate that there is no evidence for scale economies for large banks. Scale economies exist only in small and representative banks in integrated banking countries. Further, there is no evidence of scale economies in countries where commercial banking activities are separate from non-banking activities. Concerning of scope economies, it is found that global diseconomies of scope exist for loan and investment banking activities for all banks, while a small of economies of scope are found for real estate and insurance for all banks in all countries.

In general, the results on the European banking market support the existence of scale economies, though there is no consensus as to what bank size these economies are exhausted. Few studies examine the existence of economies of scope and report conflicting results. UK and Italian studies find no such evidence, whereas studies on the French banking market find evidence for economies of scope for small and medium size banks.

3.5.4.3 Methodological Issues

A portion of the recent banking literature concentrates on methodological issues. These studies attempt to examine either the robustness of the results under different methodological specification, or to utilise new methodological tools in an attempt to obtain more accurate estimates.

Given the difficulties in defining what constitutes bank's input and output, a portion of the literature examines the sensitivity of the results to different input and output specification. Nevertheless, the results from the literature are mixed. Nathan and Neave (1992) use the value-added approach and find scale efficiencies and some inefficiencies in the joint production of loan and deposits. With the application of the intermediation approach the results suggest that the production of earning assets exhibits scale inefficiencies and efficiencies in joint production. On the other hand, Favero and Papi (1993) find that the results are robust to modifications in the specification of inputs and outputs.

Another issue that has been examined in the literature is the reliability of the ordinary translog function. The major caveat of the translog function is when one output becomes zero, since in such a case the cost function becomes undefined. Many studies replace the zero value of the output with a very small number, i.e. 0.001. Gilligan and Smirlock (1984) use the value 0.001, whereas Allen and Rai (1993) use the value 0.00015. Kim (1986) and Mester (1987) select 10 percent of the mean the minimum values, respectively. Noulas *et al.* (1993) use several values (0.0001, 0.001, 0.1) in an attempt to examine the sensitivity of the results. Their sensitivity analysis indicates that the findings are robust and that scope estimates change in size and magnitude considerably. Their results indicate that banks between \$1 up to \$6 billion of assets are cost effective; larger banks do not appear to share this advantage.

Other studies examine the reliability of the translog function by estimating different cost function specifications. McKillop, Glass and Morikawa (1996) estimate four cost functions: a composite, a generalised translog, a standard translog, and a separable quadratic cost function, so as to compare the results in terms of statistical fit, regularity conditions, and measurement of scale and scope economies. They used data from five large Japanese banks over the period 1978 to 1991. The results suggest that the generalised translog and composite model mode provide equally reasonable results. This is further reinforced by the estimates of product specific scope economies and cost complementarities. In general, results indicate neither economies of scale, nor economies of scope.

McAllister and McManus (1993) suggest that a single translog cost function, produce contradictory results. To overcome this problem, they use a non-parametric regression technique to provide a global approximation to the unknown true cost function. Their findings suggest that small banks exhibit scale inefficiencies, full scale efficiencies are reached at $500 million in assets, approximately constant average cost afterwards up to $10 billion in assets.

Mester (1996) includ two additional variables in the cost function apart from input prices and outputs. The first one is the average volume of non-performing loans, which serves as a proxy for capturing the quality of the banks' output and, the financial capital of each bank which represents individual bank risk. In contrast to her previous estimates, there is evidence that banks included in the sample operate at constant returns to scale. In addition, no evidence is found for either economies or diseconomies of scope.

All previously mentioned studies consider the production cost of a bank's to be equal to its total cost. Nevertheless, Clark (1996) argues that the total cost of a bank can not be measured by the production cost alone, but opportunity cost should also be added. Accordingly, total cost comprises a bank's operating and interest expenses (production cost), and a measure of the opportunity cost of capital[48].

[48] The opportunity cost of capital is calculated by use of the capital asset pricing model: $OC_t = [r_{ft} + \beta(r_{mt} - r_{ft})] * MVE_{t-1}$, where r_{ft} denotes the risk free rate of interest, which in the current case was assumed to be the average ten-year Treasury bond rate, r_m accounts for the return on the market portfolio for the time period t, and β_b, (beta), captures the sensitivity of the security's returns to overall market movements. Beta is estimated by regressing return data for the stock against a measure of market return. The last variable, MVE, represents the market value of a bank's equity.

3.5.5 A summary of the evidence

This section surveys the empirical literature on economies of scale and scope in the banking markets. The bulk of these studies examine the cost structure of US banking industry by use of data from small banks (FCA data). Results suggest that economies of scale do not appear at more than $100 million of deposits. Later studies that utilise data from larger institutions find that such economies appear for these institutions as well. European studies appear to suggest that there are economies of scale, but there is no consensus as to the level of output at which these economies are exhausted.

Next, economies of scope, which have been examined by a few of papers do not reach any definite conclusions. US studies that concentrate on small institutions find some evidence of product complementarities. On the other hand, studies on larger institutions find no such evidence. Very few European studies examine the existence of economies of scope, and again they report conflicting results. For example, UK as well as Italian studies find no such evidence, whereas studies on the French banking market find evidence for economies of scope for small- and medium-size banks.

More recent studies address different methodological issues in an attempt to obtain better and more accurate results, noting that standard methodologies employed in the cost literature have several limitations. Finally, most studies concentrate in the US banking market. However, due to its nature (branch, state-wide, and unit banks), US is a totally different market than most of banking industries thought the world[49]. Hence, any conclusion drawn from these studies does not necessarily apply to any other market. In the current study, with the use of a new data set (see section 3.7), we test the existence of economies of scale as well as cost complementarities, in the UK banking market.

[49] For example it would take over 2,000 banking organisations to account for 90% of deposits in the US industry, while in most other developed countries 90% of deposits would be accounted for by fewer than 10 organisations.

3.6 Methodology

In the current study we utilise the translog flexible functional form to estimate the bank's cost function [Mester (1987, 1992, 1993), Fields *et al.* (1993), Jagtiani *et al.* (1995), Jagtiani and Khanthavit (1996), Rogers (1998a)]. The translog cost function for a given bank p at a time t is specified as follows:

$$\ln TC = \alpha_o + \sum_{i=1}^{n} \alpha_i \ln w_{ipt} + \sum_{k=1}^{m} \beta_k Y_{kpt} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \gamma_{ij} \ln w_{ipt} \ln w_{jpt} + \frac{1}{2} \sum_{k=1}^{m} \sum_{\varsigma=1}^{m} \delta_{k\varsigma} \ln Y_{kpt} \ln Y_{\varsigma pt} + \sum_{i=1}^{n} \sum_{k=1}^{m} z_{ik} \ln w_{ipt} \ln Y_{kpt} + \varepsilon \qquad (3.24)$$

where:

w_i = price of inputs,

Y_k = outputs,

TC = total cost

ε, ε_i = normal disturbance term identically distributed, with $\varepsilon \sim N[0, E(\varepsilon^2)]$, and $\varepsilon_i \sim N[0, E(\varepsilon_i^2)]$, and contemporaneous correlation $E(\varepsilon_t, \varepsilon_{it}) \neq 0$.

Assuming cost minimising behaviour, the logarithmic differentiation of the translog cost function equation with respect to input prices (3.24), using Shephard's lemma [Shephard (1953, 1970), Cristensen *et al.* (1973)], produces a set of equilibrium cost share equations, each for every output i. The share equations are given by:

$$S_i = \frac{\partial \ln TC}{\partial \ln w_i} = a_i + \sum_{j=1}^{n} \gamma_{ji} \ln w_{jpt} + \sum_{k=1}^{m} z_{ki} \ln Y_{kipt} + \varepsilon_i \qquad (3.25)$$

Joint estimates of the system of equations, composed of the cost equation and cost share equations produces more efficient estimates than the cost equation alone. The system of equation increases the efficiency of the estimates only if the residuals of the cost and cost share equation are contemporaneously correlated [Zellner (1962)]. Since the cost shares sum to unity, regression errors ε_i of the cost-share equations are not independent. To avoid singularity of the variance covariance matrix, a share equation must be omitted in the estimation procedures [Jagtiani and Khanthavit (1996)]. The results are invariant to the equation omitted [Diewert (1982)].

For the production and cost functions to be fully integrated, the translog cost function (3.24) must satisfy certain conditions. Regularity conditions require that the cost function is monotonically non-decreasing in input prices and outputs and concave in input prices.[50] Symmetry and homogeneity restrictions must be imposed prior to estimation to ensure the monotonicity conditions [Jorgenson (1968)]. The homogeneity conditions in input prices imply that:

$$\sum_{i=1}^{n}\alpha_i = 1 \qquad (3.26a)$$

$$\sum_{i=1}^{n}\gamma_{ij} = 0 \text{ , for all } j \qquad (3.26b)$$

$$\sum_{i=1}^{n}z_{ik} = 0\text{, for all } i \qquad (3.26c)$$

Symmetry conditions require that second-order output and input parameters satisfy:

$$\gamma_{ij} = \gamma_{ji}\text{, for all } i, j \qquad (3.27a)$$

$$z_{ik} = z_{ki}\text{, for all } k,\ i \qquad (3.27b)$$

The translog cost equation (3.24) and a cost share equation (3.25) are estimated as a system of equations, using an iterative seemingly unrelated regression technique [Zellner (1962)].

[50] The concavity of the cost function with respect to w_i can be tested for each point through the elasticity of substitution matrix. If the matrix is negative-semidefinite at every data point, then the Hessian matrix will be negative-semidefinite and the cost function will be concave at every point. Monotonicity condition is based on the notion that $S_i > 0$. If $S_i < 0$ then $\delta TC/ \delta w_i < 0$, and this in turn violates the aforementioned condition. [Fields *et al.* (1993)].

3.6.1 Overall economies of scale

As explained in detail in section 3.3, overall economies of scale can be expressed by the elasticity of cost with respect to output. Due to the multi-product nature of banks, the analysis and the interpretation of returns to scale is to some extent complex, because the average costs are only defined for single product process.

Following Mester (1987, 1992), Allen and Rai (1993), Fields *et al.* (1993), Jagtiani *et al.* (1995) and Jagtiani and Khanthavit (1996) we evaluate economies of scale as the sum of the first-order partial derivatives of the estimated cost function, with respect to output. That is:

$$SE = \sum_{k=1}^{m} \frac{\partial \ln TC}{\partial \ln Y_k} = \sum_{k=1}^{m} \left(\beta_k + \sum_{\varsigma=1}^{m} \delta_{k\varsigma} \ln Y_\varsigma + \sum_{i=1}^{n} z_{ki} \ln w_i \right) \tag{3.28}$$

where:

SE = denotes overall economies of scale.

In a competitive industry, at long-run equilibrium values of SE ≤ 1 indicate efficiency in the scale of operations corresponding to minimum costs or decreasing ray average costs. On the other hand, diseconomies of scale, SE > 1, correspond to increasing ray average costs, and indicate inefficiencies in the scale of operations.

3.6.2 Cost complementarities

As mentioned in section 3.3, economies of scope between two outputs exist when a twice differentiable multi-product cost function exhibits cost complementarities between the products. Baumol *et al.* (1988) illustrates that a sufficient, but not essential condition for overall economies of scope, is the presence of cost complementarities between output.

Cost complementarities in joint production of any two outputs, k and ζ ($COMP_{k\zeta}$), are estimated as the second-order partial derivatives of the cost function with respect to outputs. Cost complementarity exists if the marginal cost of producing one output decreases when the production of the other increases. This measure of scope economies was chosen as opposed to the other measures, because it is a local measure; Roller (1990) has shown that the translog functional form performs well locally. Several studies use the global measure of economies of scope proposed by Mester (1987). Following Mester (1987), Allen and Rai (1993), Jagtiani *et al.* (1995), and Rogers (1998a), cost complementarities can be represented as follows:

$$COMP_{k\varsigma} = \frac{\partial^2 TC}{\partial Y_k \partial Y_\varsigma} \approx$$
$$\left[\delta_{k\varsigma}\left(\beta_k + \sum_{\zeta=1}^{m}\delta_{k\varsigma}\ln Y_\varsigma + \sum_{i=1}^{n} z_{ik}\ln w_i\right) * \left(\beta_\zeta + \sum_{k=1}^{m}\delta_{k\varsigma}\ln Y_k + \sum_{i=1}^{n} z_{\varsigma i}\ln w_i\right)\right]$$

(3.29)

Cost advantages in joint production result from cost complementarities, i.e. falling marginal costs of one output associated with an increase in the level of another output. When $COMP_{k\zeta} \leq 0$, an increase in the level of output ζ reduces the marginal cost of producing output k, and there is cost complementarity between k and ζ. Operating efficiency is characterised by $COMP_{k\zeta} \leq 0$, which implies constant or decreasing marginal costs. Cost inefficiencies in joint production are indicated by $COMP_{k\zeta} > 0$.

However, in such a case one must assume that the cost structures of single and multi-product firms are comparable. Another problem arises at zero output level: as mentioned in section 3.2, translog cost function is undefined for zero output. Many studies replace the zero value of the output with a very small number, i.e. 0.001. Gilligan and Smirlock (1984) use the value 0.001, whereas Allen and Rai (1993) use the value 0.00015. Kim (1986) and Mester (1987) select 10 percent of the mean the minimum values, respectively. Consistently, in the current study, we replace the zero value with 0.0015.

To conclude both measures are based on the estimated cost frontier and, therefore, indicate whether a bank minimising cost of producing a particular output bundle could lower costs proportionately by choosing another level of output, or by changing its output mix.

3.7 Data description

One of main reasons for the limited number of studies based on UK and European banking markets in general relates to the data availability problem. Studies in the UK concentrate on building societies (see section 3.5.4.2).

The time span of the present data set is 1981 to 1995; this time interval is chosen because all major changes in the banking industry, at national and international level, occurred during this period. As mentioned in Chapter 1, British banking started changing radically with the beginning of 1980s. During that decade, banks experienced deregulation, internalisation of financial services, improvements in technology, and the path towards the single markets. During 1990s, national implementation of the EU Banking Directives came forth, whereas banks started to adjust to the new environment.

In order to provide a comprehensive treatment of the industry and to determine whether the results are stable over time, the data set has to be uniform over time. Thus, the sample covers all UK established banks that begun business on or before 1981, and were still in operation by the end of 1995. Only 78 banks are found to be in operation for all the selected time period. To attain such information, the Bank of England's Banking Act Reports were utilised. This annual publication presents a list with the institutions that are authorised by the Bank of England to accept deposits in the UK, or each year. However, from a sample of 78 the final data set comprises only 60 banks. The names of the banks constituting the sample employed in the current thesis are given in the Appendixes to Chapter 3, Table 3-25. (section 3.11.4.). No information concerning the rest twelve banks could be found.[51] After gathering all the relevant information, the data was collected from the Income Statements and Balance Sheets of each individual bank for a period of fifteen years, from 1981 to 1995. Appendix to Chapter 3 provides a description of the data for each year of the 1981-1995 period (Tables 3-6 to 3-20).

[51] Each individual bank was contacted but only four banks provided the requested information. Next we contacted the Companies House, which is the institution from which we obtain the raw data. The institution provided us with a microfiche copy for each company file. Hence, for each year and for each individual bank, the information was retrieved manually to create the data set employed in the current study.

As already mentioned, to provide a comprehensive treatment of the industry and to determine whether the results are stable over time the data employed is uniform. Nevertheless, the are certain disadvantages by using a consistent data sample. Because the time span is quiet extensive (1981-19950, a large amount of institutions that have been established after 1981, are excluded from the sample. Thus, with the establishment date restriction the data sample has been reduced to some extend. The small sample is on of the limitations of the current study. Additionally, the inclusion of all institution in operation for each year might produce different estimates. Still, these disadvantages do not outweigh the benefits of a consistent data sample. What we are trying to assess in the current study is the evolution of the UK banking system. By changing the data set from year to year one would get consistent results. Furthermore, UK banking is characterised by oligopolistic behaviour, as 70-80% of total market deposits is concentrated in three to four large banks (see Chapter 5). Hence, any additions/deletions in the number of institutions is unlikely to produce significant changes.

3.8 <u>Bank input and output specification</u>

The literature has exposed considerable difficulties in the definition and measurement of input and output in banking, since banks are multi-product firms producing joint or independent services rather than physical products. As a result, economist disagree on the correct definition of output and input in the banking industry, because each definition embeds a particular set of banking concepts. A detail analysis on the definitional problems of banks' input and output, together with the different approaches employed in the literature was given in Chapter 2, section 3.

We employed two different approaches to assess the empirical validity of the competing views concerning the classification of inputs and outputs, and to examine the sensitivity of the results to different model specification. The first one, the *intermediation approach* considers banks as intermediators of financial services, and is chosen based on its advantages over the production approach [Mester (1992), Bauer *et al.* (1993), Fields *et al.* (1993), Kaparakis *et al.* (1994), Jagtiani and Khanthavit

(1996)]. First, it is more inclusive of total banking cost, as it does not exclude interest expenses. The inclusion of both categories of costs may be more appropriate as profit maximising banks minimise total cost. Furthermore, according to Berger (1993) any exclusion of interest cost may result to bias towards finding scale economies, because banks tend to substitute interest-cost intensive purchased funds for operating cost intensive produced deposits as scale increases. In general, interest costs constitute a substantial portion of banks total cost and their exclusion may distort the empirical results.

The second, is the *value-added approach* [Nathan and Neave (1992), Berger *et al.* (1993), Berger (1993), McAllister and McManus (1993), Jagtiani *et al.* (1995), Clark (1996)]. The difference of this method is that it considers all liabilities and assets categories having some output characteristics, rather than distinguishing inputs and outputs in a mutually exclusive way. In this approach the identification of activities as inputs and outputs is based on the share of value added, which means that items of balance sheet with substantial share of value are considered important outputs. Still total costs include both operating costs as well as interest costs. Two models are estimated (see eq. 3.24), *Model I* based on the intermediation approach, and *Model II* based on the value-added approach.

Next, we proceed with the choice of the appropriate inputs and outputs. Because data availability changes intertemporally due to differences in accounting requirements, we have to include variables that are consistent for all banks for the whole sample period, to ensure that results are comparable. Moreover, most US studies distinguish between different types of deposits, i.e. demand, time and savings, and CDs, and utilise some of them as output and the others as inputs (purchased funds) [Berger and Humphrey (1990), Jagtiani and Khanthavit (1996)]. However, in our case, this cannot be applied as only the total amount of deposits is disclosed. Hence, consistent with Berg (1991) and Berg and Kim (1991), who argue that purchased funds do not qualify as an output because they do not use real resources, we utilise the total amount of deposit as an output.

Specifically, total cost (TC) is proxied by the sum of all operating costs incurred by each individual bank in the production of output and services together with interest payable. The outputs used under *Model I,* all measured in sterling pounds, are Total Deposits (y_1), and Total Loans (y_2). As Roger (1998) states, in recent studies deposits and loans are used to measure output. As far as *Model II* is concerned, a third output is included, Security Investments (y_3) (see Table 2-1 for information about input and output measurement in the literature).

In both models, inputs selected are labour (x_1), and physical capital (x_2). Labour is measured by the total number of employees at the end of the period, while physical capital is the net book value of total fixed assets. The price of labour (w_1) is derived by taking the total expenditure on employees divided by the number of employees. A proxy for the price of capital (w_2) is calculated by dividing the total expenditure on fixed assets with the net book value of fixed assets.

3.9 Empirical Results

The cost function in equation (3.24) and one of the two cost share equations are estimated as a system of equations employing an Iterative Seemingly Unrelated Regression technique imposing the linear homogeneity (3.26) and symmetry (3.27) restrictions. Since we have two inputs, labour and capital, we also have two share equations. Nevertheless, because cost shares equations sum to unity, the capital cost share equation is omitted to avoid a singular covariance residual matrix. All the estimation procedures in the current chapter and in the rest of this thesis have been obtained with the use of the LIMDEP econometric software.

An advantage of the translog function over other models used in the literature is that it allows homogeneity of degree one.[52] However, one limitation of the translog function is its inability to handle zero values. The translog requires zero costs when any output/input level is zero. In our study, due to the presence of banks with different product mixes, this constitutes a problem. To overcome this, the zero output was replaced with the value of 0.0015, following Gilligan and Smirlock (1984), Allen and Rai (1993).

[52] See Jagtianin and Khanthavit (1996).

Two models are estimated (see eq. 3.24): *Model I*, based on the intermediation approach, according to which banks are considered as intermediators of financial services, and *Model II*, based on the value-added approach, where the identification of input and output is based on the share of value added. Both models are estimated in the same way: a maximum likelihood procedure was used to estimate the parameters of the system of equations for each of the fifteen years. Parameter estimates for the each year's translog cost function for both models are shown in Appendix to Chapter 3, (Tables 3-21 and 3-22).

By looking at Tables 3-21 to 3-22 we observe that some parameters vary over time. A possible explanation may be changes in the business environment as well as changes in banks' initiated adjustments in bank costs and output (see Chapter 1, section 1.2. Market developments). For example, during the late 1980s and early 1990s due to worsening economic conditions and increasing market competition, banks cut down on loans and place emphasis on cost controls and efficiency measures. The result of this re-structuring was a sharp reduction in the number of employees. The only factor that increased was the continuously heavy investments in information technology. This gives us an indication as to why the w_2, the price of capital, varies from –1.3 to +1.6., as the major alterations in the value and the sign in the coefficients of w_2 occurred during this period.

Further to the above we estimate the cost function (eq. 3.24) for all the years (Model I) using Ordinary Least Squares to perform diagnostic tests. The estimated parameters together with diagnostic test are given in the Appendix to Chapter, section 3.11.3. As can be seen from Table 3-23 in some equations the normality hypothesis is rejected. Given that the data set it is cross-section and consists of bank of different size, bring about a possible explanation this.

To correct it we re-run the regressions with a reduced sample, after taking into accounts the potential outliers in the data sample. In all cases the normality problem is eliminated, and most importantly the estimated parameters do not change significantly (see Appendix to Chapter 3, Table 3-25). Thus, the results are fairly robust to the sample employed.

In an attempt to examine which model produces better estimates, we check the R^2s of the models. As can be seen (Tables 3-21 and 3-22), the values of R^2's for both models are in most cases above 0.90, and the difference between the two is minimal. Consistent with the high R^2's, most of the explanatory variables found to be significant for both models for all years. Consequently, both models appear as potential candidates for the measurement of scale and scope economies.

Table 3-3 presents the estimates for overall economies of scale, and the approximated standard errors of the coefficients. Overall economies of scale explain the behaviour of cost as output increases or decreases for a given bundle of outputs. The measure can be estimated by evaluating equation (3.28), using the estimated parameters from *Model I* and *Model II*, to examine how changes in scale affect the total cost as this measure is based on the estimated cost frontier it indicates whether a bank minimising the production cost of a particular output could lower costs proportionately by choosing another level of output.

In a competitive industry values of scale economies lower than one indicate efficiency in the scale of operations, brought about by minimum costs or decreasing ray average costs. Ray average cost measures the average cost of production associated with proportionate increases of all outputs as measured along a ray through the origin in the output space. On the other hand, a value greater than one indicates increasing ray average cost, which in turn suggests inefficiencies in the scale of operations (diseconomies of scale).[53]

[53] The test statistic for the economies of scale is linear and has a student-t distribution. The cost complementarities are non-linear, hence the t-statistics are asymptotically normal [Theil (1971), p.373.].

Table 3-3: Overall Economies of Scales.

Year	**Model I**	**Model II**
1995	1.2027 (0.030)*	1.2671 (0.029)*
1994	1.5686 (0.027)*	1.3909 (0.485)*
1993	1.1672 (0.0570)*	1.1887 (0.0361)*
1992	0.7620 (0.037)*	0.6344 (0.039)*
1991	1.2534 (0.059)*	1.4571 (0.055)*
1990	1.1867 (0.063)*	1.0905 (0.069)*
1989	1.3002 (0.034)*	1.4962 (0.040)*
1988	1.2174 (0.032)*	1.4363 (0.037)*
1987	1.4174 (0.044)*	1.3214 (0.053)*
1986	1.3789 (0.044)*	1.1918 (0.061)*
1985	1.2751 (0.046)*	1.1606 (0.058)
1984	1.1385 (0.057)*	1.1836 (0.052)*
1983	1.2903 (0.045)*	1.3023 (0.047)*
1982	1.2798 (0.045)*	1.2777 (0.050)*
1981	1.1358 (0.066)*	1.1738 (0.052)*

Notes: * denotes significance at 5% level.
Values of standard error in parentheses.

From Table 3-3 it is clear that the results are invariant to input and output specification, as both models yield consistent estimates. For all the periods from 1981 to 1995 (apart from 1992) the OSE (Overall Scale Economies) coefficient is greater than one and statistically significant. This suggests the existence of diseconomies of scope, i.e. banks are operating on average at greater than minimum efficiently. A possible reason might be the increasing pattern of concentration due to deregulation, which eventuated in the creation of large financial institutions; yet, the increase in the volume of activities resulted in the inefficient use of some or all inputs.

On the other hand, in attempt to become more competitive during the 1990s, made heavy investments in information technology. This steamed from the desire for long-term benefits and from increased automation of labour intensive activities (credit assessments, data processing and storage of customer information). Hence, banks may have not yet realised the benefits from deregulation and liberalisation of financial services, since the re-adjustment costs are high and have not been fully depreciated. Still, these results seem a bit surprising because most of the previous literature finds at least on some level economies of scales (Martin and Sassenou (1992), Dietch (1993), McKillop and Glass (1994)). The results are in line with the findings of Jagtiani and Khanthavit (1996) who suggest that with deregulation banks became too large to be efficient, while Mester (1992) indicates that banks are operating at slightly than minimum efficient after deregulation. Also the results are consistent with Glass and McKillop (1992) who find diseconomies of scope in the Irish banking market.

In an attempt to identify the source of these diseconomies of scales, we perform second stage regressions via OLS. Firstly, as scale economies measures indicate a more efficient use of some inputs with an increasing volume of output, an increase in size could allow more efficient organisation of resources. Thus, the first test concerns the bank size, measured by total assets [Mester (1996), Miller and Noulas (1996)].

Further, another determinant of economies of scale is market structure. According to the theory, industries that exhibit substantial scale economies may end up in an oligopolistic structure in the long run. The diseconomies of scale found here imply that the market consists of a large number of firms in the market, hence it is closer to perfect competition. Consequently, the next variable to be tested in relation to scale economies is the market concentration together with the market share of each firm. To estimate the degree of market concentration, the concentration ratio has been applied. We employ the three-bank concentration ratio, defined as the ratio of total deposits of the three largest banks to the total deposits of the banks in the UK. To measure the market share of each bank, we calculate the proportion of each individual bank's deposit in the market relative to the total market deposit.

The results obtained are reported in Table 3-4.

Table 3-4: Results from a regression of scale economies on concentration, market share, and total assets.

Variable - Year	Concentration	Market Share	Log of Total Assets
1995	2.508 (0.44)*	-1.287 (0.72)*	-0.012 (0.019)
1994	3.404 (0.77)*	0.367 (1.39)	-0.045 (0.03)
1993	2.651 (0.47)*	-1.255 (0.78)	-0.033 (0.02)
1992	1.230 (0.56)*	0.397 (0.89)	-0.013 (0.03)
1991	0.882 (0.65)	-0.238 (0.16)	0.061 (0.03)
1990	3.366 (0.80)*	-1.792 (1.35)	-0.010 (0.04)
1989	2.426 (0.51)*	0.445 (0.84)	-0.032 (0.03)
1988	1.512 (0.47)*	-0.461 (0.79)	0.016 (0.02)
1987	1.484 (0.68)*	-0.358 (1.11)	0.098 (0.39)
1986	2.508 (0.44)*	-1.287 (0.72)*	-0.012 (0.02)
1985	3.404 (0.77)*	0.3667 (1.39)	-0.045 (0.03)
1984	2.651 (0.47)*	-1.255 (0.783)	-0.034 (0.02)
1983	1.230 (0.56)*	0.397 (0.89)	-0.013 (0.03)
1982	0.882 (0.65)	-0.238 (0.16)	0.061 (0.03)
1981	3.366 (0.80)*	-1.792 (1.35)	-0.100 (0.04)

Notes: * denotes significance at 5% level.
Standard errors are in parentheses.

The negative signs of market share and total assets suggest an opposite relationship between the variables and diseconomies of scale. As mentioned above, industries that exhibit substantial scale economies may end up in an oligopolistic structure in the long run, thus fewer firms with larger market share. Nevertheless, the variable is statistically significant only in two out of the fifteen equations, whereas the size variable is always insignificant. On the other hand, concentration is positively significant throughout the sample period. The results suggests that the higher the concentration the higher the diseconomies of scale. Accordingly, the data supports the view that an oligopolistic market structure result in a less efficient organisation of resources, and thus in decreasing returns to scale. This is consistent with the 'quiet life' hypothesis; when banks enjoy greater market power and concentration, inefficiency follows not because of non-competitive pricing but because of a relaxed environment that produces no incentive to minimise cost [(Hicks 1935) 'Quiet Life Hypothesis']. In line with the aforementioned, the results suggest that the UK banks are not operating at most cost-efficient size.

Table 3-5 presents the estimates cost complementarities and the approximated standard errors of the coefficients. The measure can be estimated by evaluating equation (3.29) using the estimated parameters from *Model I* and *Model II.* Cost complementarities exist if the marginal cost of producing one output decreases when the production of the other output increases, indicating cost advantages in the joint production. When this measure is lower than zero, suggesting decreasing marginal costs, then cost complementarities are present implying economies of scope. In turn, a positive value indicates diseconomies of scope, and a value equal to zero suggests cost independence.

Table 3-5: Pairwise Cost Complementarities.

Year – Variables	**Model I**	**Model II**		
	y_1 *and* y_2	y_1 *and* y_2	y_1 *and* y_3	y_2 *and* y_3
1995	-0.0506 (0.018)*	-0.0576 (0.002)*	-0.0607 (0.002)*	-0.4416 (0.122)*
1994	0.0034 (0.001)*	-0.0042 (0.002)*	0.0013 (0.003)	0.0071 (0.008)
1993	-0.0051 (0.002)*	-0.1499 (0.052)*	0.1307 (0.0369)*	-0.1601 (0.073)
1992	-0.0628 (0.019)*	-0.5822 (0.160)*	0.1604 (0.110)	-0.8022 (0.133)*
1991	-0.0363 (0.051)	0.0181 (0.005)*	0.02002 (0.006)*	-0.0006 (0.002)*
1990	-0.0432 (0.025)	0.0866 (0.011)*	-0.0546 (0.005)*	0.0099 (0.0005)*
1989	0.1441 (0.008)*	-0.0065 (0.0004)*	-0.0540 (0.006)*	0.0582 (0.004)*
1988	0.0177 (0.001)*	0.0385 (0.003)*	-0.0205 (0.004)*	0.0104 (0.004)*
1987	0.0061 (0.006)	0.0952 (0.008)*	0.0249 (0.004)*	0.0024 (0.0003)*
1986	0.00617 (0.0036)	0.0044 (0.001)*	-0.0014 (0.003)	-0.0088 (0.001)*
1985	0.0436 (0.001)*	0.0048 (0.0005)*	0.0680 (0.0061)*	0.0121 (0.003)*
1984	0.0085 (0.001)*	0.0147 (0.001)*	0.0687 (0.0053)*	0.0687 (0.005)*
1983	0.0642 (0.003)*	0.0509 (0.005)*	0.2284 (0.0196)*	-0.0047 (0.006)
1982	-0.4924 (0.037)*	0.0340 (0.003)*	0.0708 (0.007)*	0.0011 (0.002)*
1981	-0.0583 (0.015)*	0.0256 (0.003)*	0.0196 (0.0019)*	-0.0020 (0.001)

Notes: * denotes significance at 5% level.
Standard errors are in parentheses.

As regards cost complementarities, the results are found to be sensitive to the model specification. Particularly, for *Model I* the values of cost complementarities indicate a strong interdependence between the joint production of loans and deposits for the first two years of the sample. On the other hand, *Model II* suggests diseconomies of scope for the same period and, furthermore, the existence of inefficiency in the joint production of deposits and loans during the 1980s (with the exception of 1989). The situation changes with the new decade. During the 1990s, estimates indicate a strong interdependence in joint production of loan and deposits for both models. *Model II* implying a stronger relationship than *Model I.* Cost complementarities are also present, but to a lesser extent, between loans and investments in 1986 and from 1990 to 1995 (apart from 1994) with an increasing pattern. However, in earlier years the coefficients are positive, and in some cases statistically significant, implying diseconomies of scopes.

On the other hand, findings indicate diseconomies of scope between deposits and investments for most of the years. The combining production of deposits and investment for years 1986, 1992 and 1994 are statistically insignificant, indicating the absence of any economies or diseconomies on the current product mix. Further, from 1981-1985, and for 1987, 1991, and 1993, there exists a lack of cost complementarity between deposits and investments, indicating inefficiency in the joint production of these two outputs. For years 1988-1990, and 1995 there is a statistically significant value in favour of the joint production of deposits and investments.

Thus, although banks appear to be inefficient with respect to the scale of their operation, there exist economies in the joint production of deposits and loans, and loans and investments during the 1990s. The empirical results for joint production, are not in favour of specialisation, indicating that specialisation would result in cost inefficiencies. The implication is that the joint production of these outputs resulted in lower costs. Consistent with Rogers (1998a), who finds cost and revenue complementarities after deregulation, the existence of cost complementarities in 1990s could be a sign of UK banks becoming more efficient in the face of deregulation and increased competition. On the other hand, Jagtiani *et al.* (1995) and Mester (1996) find no evidence of scope economies, whereas Jagtiani and Khanthavit find diseconomies of scope.

With deregulation, banks are able to expand their activities in other markets. The findings suggest the benefits of expanding bank powers, with these benefits coming at the expense of other financial institutions, rather than bank customers. In turn, these results imply that expansion of bank powers into other activities may provide downward pressure on prices.

3.10 Conclusions

The bulk of the empirical literature on economies of scale and scope investigate the cost structure of US markets, and report different results depending on the size of banks used in their data sample. In general, these studies tend to conclude that large bank over-perform small ones in terms of scale efficiency. For the European market little work has been done, and although some evidence suggests the presence of scale economies, there is no agreement on the level of output at which these economies are exhausted. From the above it is clear that more research must be undertaken in the area of European cost structure.

Here, we estimate a cost function for UK banks using the translog specification in an attempt we assess scale economies and cost complementarities. The data utilised covers the period from 1981 to 1995. Two models are used: the first one that classifies banks as intermediators of financial services. The second is based on the value-added approach, where the identification of input and output is based on the share of value added. Contrary to the findings of Nathan and Neave (1992), both models produce on average similar results.

The results indicate the existence of decreasing returns to scale, implying diseconomies of scale for the UK banks throughout the time period, 1981-1995. The results are in line with the finding of Jagtiani and Khanthavit (1996), Mester (1992), and Glass and McKillop (1992). Furthermore, the results suggest lack of cost complementarities at the beginning of the data period. On the other hand, with the beginning of the new decade, consistent with Rogers (1998a), there is evidence of economies of scope in the joint production. These results are somewhat different from previous findings in other countries, which, in turn, suggests that national markets and industries do not operate identically.

As noted at the beginning of this chapter, information obtained from the estimation of scale economies can be used to inform government policy by assessing the effect of deregulation on efficiency. Due to the existence of diseconomies of scale mergers and acquisitions that lead to a more concentrated financial system may imply higher cost. Thus, government policy should be designed to control the growth of banks, as this growth leads to operational inefficiencies in the form of diseconomies of scale. On the other hand the existence of cost complementarities in the joint production of loans and deposits, and loans and investments, during the 1990s suggests that deregulation benefits banks by increasing their power into new activities. This benefit, in turn, passes to the consumer in the form of lower prices, as competition increases between banks and other financial institutions.

Nevertheless, the current exploration faces a set of limitations, due to the small number of banks in the data set. Firstly, as it can be seen from section 3.5, most of the literature employed a vast amount of institutions, which can be divided sub-categories according to size (small, medium and large banks). Although the current data set consists of banks of different size, diversification cannot be performed due to the smaller number of banks included. And last, but not least, data availability changed from year to year due to differences in accounting requirements, which hunted substantially the number of potential variables for the period under consideration.

3.11 Appendix to Chapter 3

3.11.1 Data description

The following Tables give a description of basic statistics for the data for 1981-95 period.

Table 3-6: Data description for1995.

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	1,033,200	2,612,700	3.605	15.864	2,703	13,240,000
Inputs						
Labour (x_1)	208,280	566,840	3.742	16.772	425	2,907,000
Capital (x_2)	22,957	63,186	3.509	14.956	0.00150	332,000
Outputs						
Price of Labour (w_1)	46.396	27.056	1.264	4.831	8.991	138
Price of Capital (w_2)	0.23905	0.22883	1.912	7.616	0.00150	1.227
Outputs						
Deposits (y_1)	10,290,000	26,712,000	3.548	15.051	40,510	135,000,000
Loans (y_2)	9,143,400	23,610,000	3.463	14.369	36,120	116,000,000
Investments (y_3)	991,610	2,408,100	3.937	19.897	0.00150	14,630,000

Table 3-7: Data description for 1994.

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	824,880	2,273,500	3.886	17.84	2,479	11,680,000
Inputs						
Labour (x_1)	190,980	536,570	3.837	17.421	417	2,763,000
Capital (x_2)	21,122	58,030	3.342	13.334	0.00150	290,000
Price of Inputs						
Price of Labour (w_1)	41.757	23.058	0.751	3.318	0.02567	115.1
Price of Capital (w_2)	0.24227	0.22254	1.61	5.447	0.00150	0.9941
Outputs						
Deposits (y_1)	9,248,000	24,923,000	3.806	17.212	25,710	128,000,000
Labour (y_2)	8,221,600	22,479,000	3.831	17.489	19,600	117,000,000
Investments (y_3)	675,280	1,681,600	3.909	17.944	0.00150	9,250,000

Table 3-8: Data description for 1993

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	923,840	2,572,000	3.981	18.764	2571	14,140,000
Inputs						
Labour (x_1)	184,520	528,340	3.844	17.398	379	2,728,000
Capital (x_2)	21541	61122	3.648	15.916	0.00150	311,000
Price of Inputs						
Price of Labour (w_1)	41.361	28.292	1.97	8.13	0.04381	164
Price of Capital (w_2)	0.2815	0.25395	1.38	4.749	0.00150	1.135
Outputs						
Deposits (y_1)	8,942,400	25,207,000	3.925	18.095	16,530	133,000,000
Loans (y_2)	7,979,200	22,232,000	3.917	18.096	26,140	117,000,000
Investments (y_3)	621,950	1,377,770	3.563	15.560	0.00150	7,378,000

Table 3-9: Data description for1992.

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	1,082,900	3,008,500	3.938	18.333	3,892	16,270,000
Inputs						
Labour(x_1)	236,140	629,770	3.045	11.279	370	2,864,000
Capital (x_2)	19696	55573	3.702	16.478	0.00150	297,000
Price of Inputs						
Price of Labour (w_1)	39.485	28.604	2.069	8.559	0.00573	165.30
Price of Capital (w_2)	0.23529	0.21056	1.604	5.935	0.00150	1.016
Outputs						
Deposits (y_1)	8,625,400	24,578,000	4.025	18.944	23,720	129,000,000
Loans (y_2)	7,690,400	22,213,000	4.016	18.857	32,340	115,000,000
Investments(y_3)	492,400	1,183,100	3.974	20.263	0.00150	7,252,000

Table 3-10: Data description for1991.

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	1,273,454	4,002,785	4.525	21.681	4,779	23,987,000
Inputs						
Labour(x_1)	141,051	381,142	3.453	11.336	440	1,887,000
Capital (x_2)	40,512	139,718	4.209	17.701	0.0015	780,000
Price of Inputs						
Price of Labour (w_1)	35.392	20.2404	1.064	0.891	8.991052	102.502
Price of Capital (w_2)	0.2529	0.2433	2.167	7.116	0.0015	1.387
Outputs						
Deposits (y_1)	24,282,574	143,495,072	7.594	58.340	12,439	111,000,000
Loans (y_2)	6,126,727	18,573,952	4.1783	17.976	4,588	97,749,000
Investments(y_3)	367,997	853,008	3.725	14.739	0.0015	4,425,000

Table 3-11: Data description for1990.

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	1,131,855	3,347,486	3.9086	15.4419	6,396	17,881,000
Inputs						
Labour(x_1)	133,078	361,303	3.413	10.769	417	1,698,000
Capital (x_2)	36,308	123,616	4.229	18.0737	0.0015	698,000
Price of Inputs						
Price of Labour (w_1)	35.4949	24.07	1.5505	3.149	8.6997	130.138
Price of Capital (w_2)	0.2268	0.1932	1.0518	0.356	0.0015	0.75999
Outputs						
Deposits (y_1)	7,531,907	20,902,263	3.882	15.372	20,440	106,000,000
Loans (y_2)	6,178,827	18,522,689	4.1075	17.2596	4,695	95,879,000
Investments(y_3)	288,625	610,169	2.9825	8.5017	0.0015	2,821,000

Table 3-12: Data description for1989.

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	1,072,155	3,090,452	3.5789	11.9301	8,272	14,325,000
Inputs						
Labour(x_1)	130,888	352,503	3.3578	10.2428	379	1,571,000
Capital (x_2)	33,744	108,887	4.0945	17.3234	0.0015	609,000
Price of Inputs						
Price of Labour (w_1)	32.7966	20.0866	1.1444	1.0715	7.5709	95.5714
Price of Capital (w_2)	0.23813	0.21310	1.0207	0.2172	0.0015	0.78552
Outputs						
Deposits (y_1)	7,370,296	20,593,286	3.845	14.929	4,445	104,000,000
Loans (y_2)	6,137,049	18,317,532	4.01299	16.3659	4,281	94,244,000
Investments(y_3)	301,709	654,719	3.4594	13.9473	0.0015	3,752,000

Table 3-13: Data description for1988.

Variable	**Mean**	**Standard Deviation**	**Skewness**	**Kurtosis**	**Minimum £'000**	**Maximum £'000**
Total Cost	701,784	1,907,213	11.302	3.502	5,414	8,658,000
Inputs						
Labour(x_1)	120,073	322,628	10.0483	3.328	370	1,421,000
Capital (x_2)	28,328	91,611	17.324	4.087	0.0015	515,000
Price of Inputs						
Price of Labour (w_1)	31.4941	19.4224	0.6149	1.0205	7.232	88.563
Price of Capital (w_2)	0.2383	0.2473	6.919	2.159	0.0015	1.3878
Outputs						
Deposits (y_1)	6,270,258	17,370,829	14.667	3.828	4,234	87,034,000
Loans (y_2)	5,321,857	15,296,937	15.246	3.882	2,944	78,179,000
Investments(y_3)	283,250	679,057	10.849	3.335	0.0015	3,287,000

Table 3-14: Data description for 1987.

Variable	**Mean**	**Standard Deviation**	**Kurtosis**	**Skewness**	**Minimum £'000**	**Maximum £'000**
Total Cost	675,393	1,809,205	10.751	3.413	6,316	8,291,000
Inputs						
Labour(x_1)	108,115	290,312	10.244	3.367	334	1,259,000
Capital (x_2)	22,637	80,637	24.686	4.839	0.0015	499,000
Price of Inputs						
Price of Labour (w_1)	29.40552	18.3027	1.485	1.219	7.484761	89.490
Price of Capital (w_2)	0.256119	0.3276	23.610	4.212	0.0015	2.256
Outputs						
Deposits (y_1)	5,771,626	15,537,110	13.841	3.713	4,699	76,354,000
Loans (y_2)	4,410,095	12,753,016	14.137	3.763	1,911	63,986,000
Investments(y_3)	286,025	604,125	8.059	2.891	0.015	2,847,000

Table 3-15: Data description for 1986.

Variable	**Mean**	**Standard Deviation**	**Kurtosis**	**Skewness**	**Minimum £'000**	**Maximum £'000**
Total Cost	658,504	1,749,048	10.518	3.392	5,696	8,097,000
Inputs						
Labour(x_1)	99,676	269,246	10.451	3.378	315	1,204,000
Capital (x_2)	20,968	75,322	22.534	4.693	0.0015	440,000
Price of Inputs						
Price of Labour (w_1)	27.439	16.980	1.912	1.241	7.008	87.599
Price of Capital (w_2)	0.200	0.164	0.586	1.022	0.0015	0.686
Outputs						
Deposits (y_1)	5,641,050	14,980,313	12.018	3.518	3,740	72,761,000
Loans (y_2)	4,304,118	12,211,619	12.407	3.571	1,544	60,459,000
Investments(y_3)	318,199	640,176	7.266	2.753	0.0015	2,955,000

Table 3-16: Data description for1985.

Variable	**Mean**	**Standard Deviation**	**Kurtosis**	**Skewness**	**Minimum £'000**	**Maximum £'000**
Total Cost	660,763	1,769,348	9.846	3.301	5,155	7,832,000
Inputs						
Labour(x_1)	83,574	230,364	12.453	3.560	290	1,194,000
Capital (x_2)	23,566	78,909	15.404	3.938	0.015	405,000
Price of Inputs						
Price of Labour (w_1)	26.289	17.276	2.347	1.405	6.592	88.434
Price of Capital (w_2)	0.227	0.196	3.130	1.555	0.0015	0.972
Outputs						
Deposits (y_1)	5,060,501	13,611,948	10.835	3.406	10,886	64,753,000
Loans (y_2)	4,037,997	11,613,907	11.191	3.454	2,954	54,349,000
Investments(y_3)	299,084	735,873	15.580	3.736	0.0015	4,272,911

Table 3-17: Data description for 1984.

Variable	**Mean**	**Standard Deviation**	**Kurtosis**	**Skewness**	**Minimum £'000**	**Maximum £'000**
Total Cost	643,010	1,823,595	10.1569	3.3819	4,914	7,659,000
Inputs						
Labour(x_1)	79,296	224,787	14.4081	3.7602	265	1,229,000
Capital (x_2)	22,892	77,529	15.2880	3.9433	0.0015	397,000
Price of Inputs						
Price of Labour (w_1)	25.438	17.501	2.4685	1.4289	6.224	89.2676923
Price of Capital (w_2)	0.262	0.372	34.9040	5.3085	0.0015	2.75396825
Outputs						
Deposits (y_1)	5,180,488	14,119,902	10.7850	3.4134	9,636	65,515,000
Loans (y_2)	4,203,380	12,150,214	11.1234	3.4572	1,071	55,987,000
Investments(y_3)	214,590	463,318	7.9770	2.9209	0.0015	2,139,000

Table 3-18: Data description for1983.

Variable	**Mean**	**Standard Deviation**	**Kurtosis**	**Skewness**	**Minimum £'000**	**Maximum £'000**
Total Cost	547,366	1,496,701	9.301	3.256	4,480	6,094,000
Inputs						
Labour(x_1)	75,399	206,941	12.318	3.511	242	1,091,000
Capital (x_2)	19,807	65,150	14.177	3.793	0.0015	338,000
Price of Inputs						
Price of Labour (w_1)	24.604	17.785	2.662	1.499	5.931727	90.1018033
Price of Capital (w_2)	0.229	0.189	1.959	1.300	0.015	0.88049793
Outputs						
Deposits (y_1)	4,275,244	11,287,578	11.120	3.412	9,182	55,248,000
Loans (y_2)	3,403,527	9,498,617	10.723	3.374	6,641	46,015,000
Investments(y_3)	176,527	402,375	7.160	2.886	0.0015	1,639,000

Table 3-19: Data description for 1982.

Variable	**Mean**	**Standard Deviation**	**Kurtosis**	**Skewness**	**Minimum £'000**	**Maximum £'000**
Total Cost	578,489	1,638,527	10.5536	3.3933	3,951	7,703,000
Inputs						
Labour(x_1)	70,214	189,806	12.6821	3.5245	220	1,018,000
Capital (x_2)	16,699	56,652	15.1200	3.9292	0.0015	300,000
Price of Inputs						
Price of Labour (w_1)	23.887	17.452	3.2570	1.5864	6.317	90.936
Price of Capital (w_2)	0.212	0.200	5.1472	1.9230	0.0015	1.071
Outputs						
Deposits (y_1)	3,808,116	10,197,783	11.5574	3.4661	8,152	50,196,000
Loans (y_2)	3,549,608	10,642,222	14.0357	3.7446	10,194	56,841,000
Investments(y_3)	148,527	355,073	8.3712	3.0442	0.0015	1,583,000

Table 3-20: Data description for 1981.

Variable	**Mean**	**Standard Deviation**	**Kurtosis**	**Skewness**	**Minimum £'000**	**Maximum £'000**
Total Cost	463,595	1,253,747	8.9726	3.2010	2,989	5,261,000
Inputs						
Labour(x_1)	63,301	163,773	9.1069	3.1528	250	744,700
Capital (x_2)	14,690	45,388	10.4232	3.4013	0.0015	201,000
Price of Inputs						
Price of Labour (w_1)	23.1540	17.8635	3.4149	1.6755	5.242	91.770
Price of Capital (w_2)	0.2407	0.2830	21.6182	3.9189	0.0015	1.938
Outputs						
Deposits (y_1)	3,168,711	8,288,995	11.5020	3.4549	5,081	39,709,000
Loans (y_2)	2,450,043	6,783,022	11.1111	3.4345	1,645	32,412,000
Investments(y_3)	118,874	287,926	8.9432	3.1442	0.0015	1,205,500

3.11.2 Parameter estimates from the traslog cost function

Table 3-21: Parameter estimates from Model I

Variable	1995	1994	1993	1992	1991	1990	1989	1988
Constant	5.682	8.956	9.300	12.644	18.215	13.630	15.709	14.595
	(3.39)*	(2.42)*	(3.12)*	(3.73)*	(3.64)*	(3.30)*	(3.69)*	(3.17)*
w_1	-2.672	-2.989	-2.836	-2.931	-3.644	-2.033	-3.071	-3.504
	(0.68)*	(0.56)*	(0.63)*	(0.69)*	(0.64)*	(0.62)*	(0.74)*	(0.64)*
w_2	0.291	0.541	-0.544	0.252	-1.310	-0.218	-0.906	-0.943
	(0.46)	(0.38)	(0.51)	(0.55)	(0.53)*	(0.50)	(0.56)	(0.47)*
Y_1	2.642	2.287	2.041	2.416	0.946	1.118	0.778	0.833
	(0.31)*	(0.22)*	(0.23)*	(0.30)*	(0.24)*	(0.25)*	(0.27)*	(0.24)*
Y_2	-1.962	-1.976	-1.897	-2.594	-2.206	-1.851	-1.642	-1.538
	(0.28)*	(0.22)*	(0.28)*	(0.36)*	(0.34)*	(0.30)*	(0.30)*	(0.26)*
$w_1.w_2$	-0.214	-0.271	-0.043	-0.126	0.462	0.168	0.310	0.231
	(0.144)	(0.10)*	(0.10)	(0.11)	(0.14)*	(0.13)	(0.14)*	(0.13)*
$Y_1.Y_2$	-0.026	-0.003	0.010	0.027	0.083	0.075	0.069	0.058
	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)*	(0.02)*	(0.03)*	(0.03)*
$w_1.Y_1$	-0.535	-0.539	-0.473	-0.488	-0.174	-0.409	-0.293	-0.279
	(0.06)*	(0.05)*	(0.05)*	(0.05)*	(0.07)*	(0.06)*	(0.07)*	(0.06)*
$w_1.Y_2$	0.866	0.902	0.855	0.867	0.818	0.736	0.791	0.842
	(0.07)*	(0.05)*	(0.05)*	(0.06)*	(0.06)*	(0.06)*	(0.07)*	(0.06)*
$w_2.Y_1$	1.060	0.290	0.298	0.655	0.621	0.370	-0.0174	-0.029
	(0.24)*	(0.18)	(0.18)*	(0.20)*	(0.19)*	(0.20)*	(0.106)	(0.09)
$w_2.Y_2$	-1.089	-0.338	-0.218	-0.675	-0.581	-0.411	0.053	0.0766
	(0.25)*	(0.19)*	(0.19)	(0.23)*	(0.22)*	(0.20)*	(0.09)	(0.08)
R^2 *(cost)*	0.93	0.95	0.95	0.95	0.92	0.95	0.96	0.96
R^2 *(share)*	0.74	0.80	0.81	0.82	0.77	0.74	0.77	0.77

Table 3-21 (continued):

Variable	1987	1986	1985	1984	1983	1982	1981
Constant	13.714	13.738	15.340	12.499	12.077	9.829	10.840
	(3.69)*	(3.67)*	(4.15)*	(3.17)*	(3.67)*	(3.56)	(3.51)*
w_1	-1.050	-1.250	-1.352	-0.802	-0.755	-1.135	-0.564
	(0.81)	(0.85)	(0.80)*	(0.70)	(0.77)	(0.76)	(0.73)
w_2	0.438	1.608	0.875	0.768	1.225	0.693	0.294
	(0.59)	(0.98)	(0.90)	(0.45)*	(0.78)	(0.60)	(0.63)
Y_1	-1.008	-0.831	-1.954	-1.429	0.297	1.647	-0.855
	(0.69)	(0.74)	(1.09)*	(0.78)*	(1.30)	(1.10)	(1.01)
Y_2	0.042	0.060	0.797	0.501	-1.159	-2.139	0.047
	(0.64)	(0.67)	(0.93)	(0.59)	(1.15)	(0.98)*	(0.83)
$w_1.w_2$	0.048	0.006	-0.091	0.061	-0.064	-0.150	-0.025
	(0.08)	(0.12)	(0.18)	(0.10)	(0.11)	(0.11)	(0.11)
$w_1.Y_1$	0.319	0.340	0.249	0.536	0.221	0.194	0.469
	(0.14)*	(0.13)*	(0.16)	(0.14)*	(0.17)	(0.17)	(0.14)*
$w_1.Y_2$	0.478	0.458	0.543	0.215	0.508	0.562	0.262
	(0.13)*	(0.12)*	(0.15)*	(0.13)	(0.16)*	(0.16)*	(0.14)*
$w_2.Y_1$	-0.070	-0.053	-0.007	-0.079	-0.020	0.066	-0.068
	(0.10)	(0.17)	(0.19)	(0.11)	(0.17)	(0.16)	(0.194)
$w_2.Y_2$	0.028	-0.069	-0.162	-0.084	-0.140	-0.156	0.025
	(0.09)	(0.17)	(0.17)	(0.10)	(0.16)	(0.15)	(0.17)
R^2 *(cost)*	0.91	0.89	0.90	0.94	0.91	0.92	0.89
R^2 *(share)*	0.73	0.72	0.75	0.78	0.73	0.74	0.68

Notes: (1) Standard errors of estimated parameters are shown in parentheses/ * significant with 95% confidence interval.
(2) Certain variables are excluded, due to high multicolinearity.

Table 3-22: Parameter estimates from Model II

Variable	1995	1994	1993	1992	1991	1990	1989	1988
Constant	7.900 (3.30)*	7.745 (2.36)*	7.652 (2.91)*	9.377 (3.64)*	13.150 (3.31)*	13.015 (3.53)*	18.642 (4.02)*	17.669 (3.63)*
w_1	-2.633 (0.65)*	-2.673 (0.55)*	-2.930 (0.63)*	-2.614 (0.702)*	-2.682 (0.59)*	-1.911 (0.59)*	-3.45 (0.63)*	-3.501 (0.59)*
w_2	0.313 (0.43)	0.438 (0.41)	-0.305 (0.43)	0.852 (0.50)*	-0.768 (0.52)	-0.240 (0.46)	-1.59 (0.47)*	-1.03 (0.39)*
Y_1	-0.535 (0.95)	2.319 (0.53)*	0.659 (0.402)	0.555 (0.57)	0.424 (0.241)*	0.406 (0.23)*	0.31 (0.28)*	0.43 (0.24)*
Y_2	1.007 (0.757)	-1.754 (0.55)*	-0.057 (0.39)	-0.254 (0.60)	-0.992 (0.23)*	-1.059 (0.27)*	-1.663 (0.31)*	-1.600 (0.27)*
Y_3	-0.103 (0.28)	-0.205 (0.23)	-0.199 (0.25)	-0.034 (0.265)	-0.011 (0.09)	0.047 (0.08)	0.229 (0.11)*	0.12 (0.09)*
$w_1.w_2$	-0.211 (0.14)	-0.248 (0.10)*	-0.090 (0.10)	-0.156 (0.104)*	0.184 (0.12)	0.050 (0.11)	0.176 (0.12)	0.111 (0.11)
$Y_1.Y_2$	0.023 (0.04)	0.008 (0.03)	0.051 (0.04)	0.050 (0.03)	0.049 (0.02)	0.072 (0.03)*	0.107 (0.05)*	0.094 (0.03)*
$Y_1.Y_3$	0.362 (0.11)*	-0.028 (0.09)*	0.143 (0.04)*	0.272 (0.08)*	0.036 (0.01)*	0.021 (0.02)	0.045 (0.02)*	0.070 (0.02)*
$Y_2.Y_3$	-0.397 (0.13)*	0.015 (0.09)	-0.210 (0.06)*	-0.334 (0.09)*	-0.020 (0.02)	-0.014 (0.019)	-0.05 (0.06)*	-0.072 (0.01)*
$w_1.Y_1$	-0.510 (0.06)*	-0.536 (0.04)*	-0.460 (0.05)*	-0.500 (0.05)*	0.092 (0.07)	0.062 (0.07)	0.048 (0.07)	-0.022 (0.06)
$w_1.Y_2$	0.657 (0.12)*	0.730 (0.11)*	0.520 (0.10)*	0.68 (0.11)*	0.364 (0.09)*	0.238 (0.10)*	0.51 (0.09)*	0.572 (0.08)*
$w_1.Y_3$	0.212 (0.10)*	0.156 (0.09)*	0.357 (0.02)*	0.177 (0.09)*	0.518 (0.08)*	0.614 (0.09)*	0.415 (0.07)*	0.358 (0.06)*
$w_2.Y_1$	0.411 (0.31)	0.293 (0.22)	0.375 (0.17)*	0.898 (0.19)*	-0.480 (0.12)*	-0.621 (0.11)*	-0.254 (0.07)*	-0.240 (0.06)*
$w_2.Y_2$	-0.450 (0.29)	-0.296 (0.21)	-0.227 (0.16)	-0.823 (0.20)*	0.015 (0.02)	0.006 (0.02)	0.011 (0.02)	-0.011 (0.02)
$w_2.Y_3$	0.023 (0.07)	-0.019 (0.07)	-0.105 (0.06)*	-0.211 (0.07)*	-0.045 (0.03)	-0.051 (0.03)*	-0.114 (0.041)	-0.071 (0.03)
R^2 *(cost)*	0.93	0.96	0.96	0.96	0.93	0.96	0.96	0.96
R^2 *(share)*	0.74	0.80	0.82	0.82	0.77	0.74	0.77	0.77

Table 3-22 (continued):

Variable	1987	1986	1985	1984	1983	1982	1981
Constant	23.098 (5.05)*	20.554 (5.58)*	15.073 (4.28)*	14.535 (3.74)*	17.52 (4.22)*	13.14 (4.84)*	14.55 (3.95)*
w_1	-2.429 (0.91)*	-1.522 (0.95)	-0.760 (0.88)	-0.806 (0.84)	-1.174 (0.83)	-1.249 (0.82)	-0.687 (0.83)
w_2	0.981 (0.62)	1.861 (0.94)*	0.556 (0.98)	0.531 (0.54)	1.446 (0.74)*	0.674 (0.61)	0.630 (0.61)
Y_1	-0.518 (0.98)	0.699 (1.22)	-0.017 (1.44)	-0.511 (1.26)	1.005 (1.72)	1.836 (1.41)	0.781 (1.23)
Y_2	-1.533 (1.005)	-2.372 (1.35)*	-1.185 (1.35)	-0.831 (1.17)	-2.68 (1.60)*	-2.81 (1.23)*	-2.189 (1.13)*
Y_3	0.334 (0.13)*	0.163 (0.15)	0.057 (0.10)	0.228 (0.13)*	0.337 (0.15)*	0.138 (0.14)	0.256 (0.13)*
$w_1.w_2$	0.007 (0.09)	0.044 (0.12)	0.399 (0.30)	0.087 (0.11)	-0.045 (0.11)	-0.140 (0.11)	-0.029 (0.11)*
$Y_1.Y_2$	0.331 (0.25)	0.531 (0.31)*	0.289 (0.47)	-0.274 (0.20)	-0.301 (0.39)	-0.548 (0.41)	-0.214 (0.21)
$Y_1.Y_3$	0.001 (0.04)	0.0565 (0.05)	0.061 (0.05)	0.032 0.05	0.015 (0.08)	0.034 (0.08)	0.051 (0.01)
$Y_2.Y_3$	-0.042 (0.03)	-0.088 (0.05)*	-0.066 (0.05)	-0.069 (0.05)	-0.066 (0.07)	-0.053 (0.07)	-0.097 (0.05)*
$w_1.Y_1$	0.461 (0.14)*	0.356 (0.14)	0.168 (0.16)	0.470 (0.14)*	0.251 (0.16)	0.210 (0.16)	0.351 (0.15)*
$w_1.Y_2$	0.452 (0.13)*	0.467 (0.12)*	0.603 (0.15)*	0.286 (0.13)*	0.526 (0.16)*	0.562 (0.16)*	0.392 (0.15)*
$w_1.Y_3$	-0.040 (0.02)*	0.001 (0.001)	0.002 (0.001)	0.002 (0.001)	0.003 (0.001)	0.002 (0.001)	0.002 (0.002)
$w_2.Y_1$	-0.174 (0.10)*	-0.204 (0.19)	-0.032 (0.19)	-0.087 (0.12)	-0.096 (0.17)	-0.048 (0.18)	-0.311 (0.18)*
$w_2.Y_2$	0.064 (0.09)	0.037 (0.16)	-0.053 (0.18)	-0.048 (0.10)	-0.081 (0.15)	-0.050 (0.15)	0.177 (0.15)
$w_2.Y_3$	0.022 (0.03)	-0.026 (0.19)	-0.367 (0.16)*	-.0166 (0.09)	0.012 (0.14)	0.058 (0.15)	0.115 (0.12)
R^2 *(cost)*	0.92	0.90	0.91	0.94	0.93	0.93	0.91
R^2 *(share)*	0.74	0.73	0.75	0.72	0.74	0.74	0.69

Notes: (1) Standard errors of estimated parameters are shown in the parentheses/ * significant with 95% confidence interval.
(2) Certain variables are excluded, due to high multicolinearity

3.11.3 OLS estimation

Table 3-23: OLS parameter estimates and diagnostic tests from Model I

Variable	1995	1994	1993	1992	1991	1990	1989	1988
Constant	6.362 (6.24)	1.975 (4.99)	2.755 (4.62)	-1.223 (4.26)	7.124 (5.01)	7.003 (3.88)*	9.211 (3.51)*	8.134 (3.13)*
w_1	-0.671 (1.43)	-0.275 (1.19)	-0.403 (1.09)	0.858 (0.90)	-1.067 (1.08)	-0.229 (0.79)	-0.997 0.830	-1.402 (0.76)*
w_2	0.154 (0.90)	0.431 (0.80)	-0.586 (0.76)	-0.705 (0.74)	-1.019 (0.82)	0.016 (0.57)	-0.411 (0.45)	-1.402 (0.76)
Y_1	-4.645 (1.17)*	-2.662 (1.23)*	-3.909 (0.98)*	-1.591 (0.74)*	0.625 (0.64)	-0.667 (0.57)	-1.359 (0.33)*	-0.914 (0.35)*
Y_2	4.814 (1.09)*	3.348 (1.49)*	4.359 (1.12)*	-1.591 (0.74)*	0.619 (0.64)	0.601 (0.55)	1.097 (0.37)*	0.840 (0.42)*
$Y_1.Y_2$	0.021 (0.02)	0.008 (0.01)	0.022 (0.01)	0.017 (0.01)	0.027 (0.01)*	0.034 (0.01)*	0.038 (0.01)*	0.028 (0.01)*
$w_1.w_2$	0.005 (0.17)	-0.021 (0.15)	-0.045 (0.14)	0.069 (0.12)	0.099 (0.14)	0.021 (0.08)	0.049 (0.07)	0.026 (0.08)
$w_1.Y_1$	1.235 (0.30)*	0.814 (0.31)*	1.105 (0.25)*	0.532 (0.18)*	0.297 (0.17)*	0.261 (0.17)	0.510 (0.09)*	0.434 (0.10)*
$w_2.Y_1$	-0.009 (0.04)	-0.023 (0.04)	0.064 (0.04)*	-	-	-	-	0.019 (0.04)
$w_1.Y_2$	-1.206 (0.27)	-0.822 (0.36)*	-1.106 (0.27)	-0.591 (0.19)*	-0.222 (0.18)	-0.262 (0.16)	-0.465 (0.10)*	-0.349 (0.12)*
$w_2.Y_2$	-	-	-	0.039 (0.38)	0.053 (0.03)	-0.005 (0.03)	0.018 (0.02)	-0.003 (0.03)
R^2	0.92	0.95	0.95	0.95	0.92	0.95	0.95	0.96
Diagnostic Tests								
Serial Correlation	0.500	0.072	1.400	0.047	1.856	0.741	0.709	0.274
Functional Form	0.008	4.663	2.143	0.208	1.967	0.023	0.047	0.085
Normality	6.081	1.430	7.472	3.411	11.221	0.300	1.384	3.934
Heterosced.	0.47E-3	0.943	5.897	3.928	0.582	1.9959	0.915	0.941

Table 3-23 (continued):

Variable	**1987**	**1986**	**1985**	**1984**	**1983**	**1982**	**1981**
Constant	8.437 (3.20)*	7.802 (3.49)*	8.338 (3.83)*	5.342 (1.96)*	5.678 (2.13)*	6.325 (2.27)*	7.131 (2.24)
w_1	-0.932 (0.79)	-1.201 (0.93)	-1.860 (0.85)*	-0.979 (0.47)*	-0.656 (0.44)	-0.556 (0.50)	-0.794 (0.49)
w_2	0.121 (0.64)	1.049 (0.73)	1.29 (0.81)*	0.551 (0.31)*	0.883 (0.44)	0.785 (0.42)*	0.253 (0.42)
Y_1	-0.531 (0.39)	-0.297 (0.73)	0.815 (0.52)	0.218 (0.26)	0.128 (0.26)	0.136 (0.29)	-0.369 (0.28)
Y_2	0.335 (0.45)	0.380 (0.47)	-0.715 (0.50)	0.014 (0.2)	0.017 (0.30)	0.109 (0.35)	0.270 (0.27)
$Y_1.Y_2$	0.029 (0.01)*	0.021 (0.01)*	0.016 (0.01)	0.019 (0.01)*	0.024 (0.01)*	0.029 (0.01)*	0.031 (0.01)*
$w_1.w_2$	0.150 (0.10)	-0.072 (0.12)	-0.112 (0.12)	-0.061 (0.06)	-0.077 (0.07)	-0.056 (0.07)	0.049 (0.07)
$w_1.Y_1$	0.206 (0.10)*	0.192 (0.11)*	-0.032 (0.14)	0.087 (0.07)	0.075 (0.07)	0.043 (0.08)	0.136 (0.08)*
$w_2.Y_1$	-0.133 (0.05)*	-0.126 (0.07)*	-0.052 (0.08)	-0.074 (0.04)*	-0.099 (0.05)*	-0.109 (0.05)*	-0.122 (0.05)*
$w_1.Y_2$	-0.123 (0.12)	-0.126 (0.12)	0.140 (0.14)	-0.029 (0.07)	-0.041 (0.08)	-0.015 (0.08)	-0.077 (0.07)
$w_2.Y_2$	0.097 (0.05)*	0.087 (0.7)	-.0189 (0.07)	0.050 (0.05)	0.054 (0.04)	0.064 (0.04)	0.092 (0.04)*
R^2	0.95	0.93	0.94	0.98	0.98	0.97	0.97
Serial Correlation	1.011	0.579	0.634	0.871	0.0106	0.55E-3	0.368
Functional Form	1.606	1.191	14.806	0.335	0.035	0.213	2.529
Normality	189.41	848.78	177.535	28.041	8.944	1.961	0.029
Heterosced.	0.547	0.339	0.084	0.254	0.181	0.772	1.349

Notes: (1) Standard errors of estimated parameters are shown in parentheses/ * significant with 95% confidence interval.
(2) Certain variables are excluded, due to high multicolinearity.

Table 3-24: Reduce sample estimates

Variable	1995	1993	1991	1987	1986	1985	1984	1983
Constant	4.969	3.764	6.720	11.081	9.163	3.156	7.935	6.627
	(9.68)	(5.27)	(8.11)	(2.40)*	(2.28)*	(7.92)	(2.22)*	(2.22)*
w_1	0.562	3.745	-1.112	-1.651	-2.539	-0.858	-2.382	-2.057
	(2.29)	(1.36)*	(1.94)	(1.12)	(0.89)*	(1.50)	(0.747)*	(0.58)*
w_2	0.476	-0.239	-0.858	-0.444	1.351	0.102	0.071	0.515
	(1.18)	(0.85)	(1.56)	(0.52)	(0.49)*	(1.02)	(0.41)	(0.71)
Y_1	-3.043	-0.502	0.111	-1.124	-0.908	0.559	-0.708	-0.465
	(1.67)*	(1.03)*	(1.02)	(0.30)*	(0.37)*	(0.83)	0.48	(0.57)
Y_2	4.076	6.787	-0.093	0.618	1.126	-0.063	0.858	0.733
	(1.28)*	(1.19)*	(1.01)	(0.41)	(0.48)*	(0.91)	(0.42)*	(0.55)
$Y_1.Y_2$	0.024	0.017	0.023	0.038	0.014	0.014	0.017	.0153
	(0.02)	(0.01)	(0.02)	(0.01)*	(0.011)	(0.02)	(0.01)*	(0.01)
$w_1.w_2$	0.024	-0.368	0.131	0.151	-0.301	-0.040	-0.102	-0.165
	(0.02)	(0.15)*	(0.25)	(0.10)	(0.10)*	(0.16)	(0.07)	(0.10)*
$w_1.Y_1$	0.816	1.362	0.115	0.442	0.585	0.003	0.549	0.476
	(0.41)*	(0.25)*	(0.26)	(0.15)*	(0.17)*	(0.18)	(0.22)*	(0.27)*
$w_2.Y_1$	0.034	0.112	0.033	-0.028	0.043	0.087	0.097	0.070
	(0.07)	(0.05)*	(0.06)	(0.76)	(0.08)	(0.15)	(0.09)	(0.11)
$w_1.Y_2$	-1.046	-1.682	-0.022	-0.324	-0.478	0.045	-0.421	-0.372
	(0.33)	(0.29)*	(0.27)	(0.13)*	(0.19)*	(0.19)	(0.21)*	(0.27)
$w_2.Y_2$	-	-	-	0.021	0.070	0.101	-0.085	-0.079
				(0.06)	(0.07)	0.166	(0.08)	(0.10)
R^2	0.90	0.96	0.94	0.99	0.99	0.95	0.99	0.99
No of obs	35	35	29	26	29	35	26	29
Diagnostic Tests								
Serial Correlation	0.115	3.327	0.501	0.021	0.154	1.286	0.052	0.101
Functional Form	0.507	0.635	3.37	0.25E-3	6.080	0.22E-3	0.810	0.658
Normality	2.648	1.498	5.34	0.015	2.705	4.3320	0.238	1.135
Heterosced.	0.018	1.094	0.094	0.3E-3	1.174	0.051	0.459	0.069

3.11.4 Banks in the sample

Table 3-23 provides a list with the names of the banks comprising the data sample.

Table 3-25: Banks in the sample

1. ANZ Grindllays
2. Alexanders Discount plc
3. Allied Trust Bank Ltd
4. Anglo Romanian Bank Ltd
5. Henry Ansbacher & Co Ltd
6. Arbutham Latham & Co Ltd
7. Bank Leumi (UK) Ltd
8. Bank of America International Ltd
9. Bank of Cyprus (London) Ltd
10. Bank of Tokyo Mitsubishi (UK) Ltd
11. Bank of Wales
12. Barclays Bank plc
13. Banque National de Paris plc
14. The British Bank of Middle East
15. The British Linean Bank Ltd
16. Brown, Shipley & Co Ltd
17. Cater Allen Ltd
18. Chartered Trust plc
19. Charterhouse Bank Ltd
20. Citibank International plc
21. Clive Discount Company Ltd
22. Close Brothers Ltd
23. Clydesdale Bank plc
24. The Co-operative Bank plc
25. Coutts & Co
26. Daiwa Europe Bank plc
27. First National Commercial Bank plc
28. Robert Fleming & Co Ltd

29.Gerrard and National Ltd

30.Guiness Mahon Ltd

31.Hambros

32.Havana International Bank Ltd

33.Hill Samuel Bank Ltd

34.Hoare & Co

35.IBJ International Bank

36.Italian International Bank

37.Leopold Joseph & Son Ltd

38.King & Saxon Ltd

39.Kleinworth Benson Ltd

40.Lazard Brothers & Co Ltd

41.Lloyds Bank plc

42.Merril Lynch International Bank Ltd

43.Midland Bank

44.Samuel Montaqu & Co Ltd

45.Moscow Norodny Bank Ltd

46.Morgan Greenfell & Co Ltd

47.National Westminster Bank Ltd

48.Rea Brothers Group

49.Royal Bank of Scotland

50.Royal Bank of Canada Europe Ltd

51.Schroders Leasing Ltd

52.Henry Schroders Wagg & Co Ltd

53.Standard Charter Bank

54.Singer and Friedlander Ltd

55.TSB Bank

56.Uster Bank Ltd

57.London Scottish Bank plc

58.Wintrust Securities Ltd

59.Yorkshire Bank plc

60.Roy Scott Trust plc

CHAPTER 4: The X-efficiency of Banking Institutions - Theory and Evidence

4.1 Introduction

In the current chapter we test the X-efficiency of UK banks by estimating an efficient frontier and by measuring the average differences between observed banks and banks at the frontier. According to Leibenstein (1966) X-efficiency accounts for differences in costs that cannot be explained by differences in scale or other observable characteristics. By focusing on scale and scope economies we can only gain limited information on efficiency, and thus we want to test more formally on X-efficiency. Three models will be applied to check the robustness of estimates to alternative specifications.

As we showed in Chapter 3, several studies have concentrated on the efficiency of banking institutions. However, until recently, only scale and scope efficiencies have been extensively studied, and relatively little attention is paid to X-inefficiencies i.e. to the differences in efficiency or deviations from the efficient frontier. The effect of differences in managerial ability to maximise revenue or minimise cost appears to be greater than any cost effects brought about by the choice of scale and scope of production. The recent literature suggests that X-inefficiencies account for 20% or more of costs in banking, whereas scale and product mix inefficiencies account for less than 5% of costs [Berger *et al.* (1993)].

Although a considerable amount of general research has taken place on X-efficiency since its introduction in the 1960s, published technical research on X-efficiency of financial institutions has only appeared in the last few years. Most studies concentrate on the US banking system, with only a handful of papers measuring the efficiency of banks outside the US. Hence, in terms of maturity and breadth the efficiency research has not kept pace with changes in the financial services industry.

This chapter is organised as follows: Section 2 presents the different approaches employed on the literature in the measurement of X-efficiency. Section 3 gives a detailed description of the literature review. In section 4 the methodology to be applied is explained. Section 5 give a description of the data used together with the determination of input and output. Section 6 reports the results obtain together with the relevant comments. Finally, section 7 summarises and presents conclusions.

4.2 Parametric and non-parametric approaches for measuring X-efficiency

In evaluating the performance of banks, the first task is to separate production units that -by some standards- perform well from those that perform poorly. This can be done through **frontier analysis,** also known called X-efficiency, which is defined as the ratio of minimum costs that could have been expended to produce a given output bundle to actual cost expended, and varies between 0 and 100 percent [Berger (1995)].

Frontier analysis covers all technical and allocative efficiencies of individual firms, that are distinct from scale and scope economies. Technically, efficiency determines the proportional decrease in input usages, which could be achieved if the firm operated on the production frontier, whereas allocative efficiency indicates the proportional decrease in cost if the right mix of input had been utilised. Essentially, this kind of efficiency is measured, or modelled, as distance measures reflecting a firm's own position in relation to its efficient production or cost frontier. The real advantage of frontier estimation is that it determines objectively numerical efficiency values and ranking of firms that would not be otherwise available [Berger and Humphrey (1997)].

Due to its advantages, frontier analysis is viewed as the benchmark to the measurement of the performance of production units. It permits to individuals with little institutional knowledge to select best practice firms within the banking industry by having actual numerical efficiency values for each firm. In the case of individuals with sufficient knowledge, the analysis allows the identification of best practice areas with complex service operation, a task not always feasible via traditional benchmark techniques, e.g. ratio analysis [Berger and Humphrey (1997)].

The major econometric problem lies in distinguishing X-efficiency differences from random errors that may temporarily appears as relatively high or low costs for certain firms. The major frontier efficiency models employed extensively in the banking X-efficiency literature are the stochastic econometric frontier approach, the thick frontier approach, the distribution-free approach, and the data envelopment analysis. Each of these approaches maintains a different set of assumptions on probability distributions of X-efficiencies and random errors to distinguish between these two explanations of cost dispersion.

4.2.1 The Stochastic Econometric Frontier Approach (SFA)

The basic stochastic econometric frontier model states that a firm's observed cost will deviate from the cost frontier because of random noise and possible inefficiencies. The key to this approach is the two-part composed error term, where one part accounts for inefficiencies and one part for statistical noise. These two components are separated by assuming that inefficiencies are drawn from a one-sided distribution (usually the half-normal) and the random fluctuations are drawn from a symmetric distribution (usually the normal). The logic behind this is that inefficiencies must have a truncated distribution because they increase costs only above frontier levels, whereas random noise can either decrease or increase costs. Both inefficiencies and the random error are assumed to be orthogonal to input, output, or environmental variables specified in the estimated model.

The major drawback of this method is that the half-normal assumption on inefficiencies is relatively inflexible and embodies the arbitrary restriction that most firms operate near full efficiency. To overcome this inflexibility some studies [Greene 1990, and Berger and De Young 1997] specified a more general truncated normal distribution or a gamma distribution. Although these methods allow for more flexibility in the assumed distribution, are closer to the symmetric normal distribution assumed for the random error, it becomes more difficult to separate inefficiencies from random errors.

Under the stochastic econometric frontier approach the cost frontier is formulated by estimating a cost function which relates observed costs to output quantities and input prices, allowing for random error and inefficiencies [see *Ainger et al.* (1977)]. This frontier can be expressed as:

$$C = C(Y, w, \varepsilon) \tag{4.1}$$

Where:
C = a measure of cost,
Y = a vector of output quantities,
w =a vector of input price, and
$\varepsilon = v + u$ = the error components.

The error component v represents the symmetric disturbance, which is assumed to be independently and identically normally distributed with zero mean and σ_v^2 variance. The error component u is assumed to be half-normally distributed, and captures inefficiencies. The economic logic behind this specification is that the production process of a firm is subject to two economically distinguishable random disturbances, with different characteristics. Each firm's cost must lie on or above its cost frontier. The non-negative disturbance u denotes any deviations from that frontier, with deviations resulting from factors under the firm's control. On the other hand, the frontier can vary across firms or across time for the same firm. Hence, with a stochastic frontier, v depends on external factors, and on measurement error.

4.2.2 Thick Frontier Approach (TFA)

The thick frontier approach compares the average efficiencies of groups of banks. Instead of estimating the actual cost or production frontier edge, the thick frontier is estimated. The thick frontier cost function is being formed by the banks in the lowest average quartile, which are assumed to be of greater than average efficient and they form the thick frontier. Consistently, banks in the high average cost quartile are more likely to operate at less than average efficiency. Accordingly, a cost function is estimated for banks in the highest average quartile, which are identified as having less than average efficiency. '*Differences in error terms within the highest and lowest* *quartiles are assumed to reflect random error, while predicted cost differences between these quartiles are assumed to reflect inefficiencies plus exogenous differences in output quantities and input prices. Banks are stratified by size class before the quartiles are formed to ensure that a board range of banks are represented in each quartile and to reduce the relationship between the quartile selection criterion and the dependent variable in the regressions.'*[54]

To decompose the predicted cost differences between the highest and lowest quartiles into market factors and inefficiency residuals, after estimating the thick frontier for each quartile several measures are calculated [Berger and Humphrey (1991)]. Following Bauer *et al.* (1992), the difference in predicted unit costs is given by:

$$Diff = \left[A\hat{C}^{Qj} - A\hat{C}^{Qi} \right] / A\hat{C}^{Qi} \qquad (4.2)$$

where:

Diff = the proportional increase in predicted unit costs of the Q_j data relative to the Q_i data evaluated at the size class means.

Q_j = the set of banks in the highest quartile,

Q_i = denotes the set of banks in the lowest quartile,

$A\hat{C}^{Qi} \equiv \hat{C}^{Qi}(X^{Qi}) / TA^{Qi}$, and denotes the predicted unit costs,

$\hat{C}^{Qi}$ = the predicted cost function using the parameters of the equations in the model, obtained when using Q_i data,

X^{Qi} = the vector of mean outputs and other regressors for the size class for the i^{th} quartile,

TA^{Qi} = the mean total assets for the size class for ith quartile,

To capture exogenous differences in markets where banks operate, this approach further assumes differences in output levels and mix, branch offices, other assets, input prices, and purchased funds levels. Consequently, the part of *Diff* owing to exogenous market factors is given by the following expression:

[54] Bauer *et al.* (1993), p.p. 388.

$$Market = \left[A\hat{C}^{Qj*} - A\hat{C}^{Qi} \right] / A\hat{C}^{Qi} \tag{4.3}$$

where $A\hat{C}^{Qj*} \equiv \hat{C}^{Qi}(X^{Qj}) / TA^{Qj}$, which is the predicted unit cost for Q_j data calculated using the efficient Q_i technology, rather than the inefficient Q_j technology. Hence, *Market* captures the effects of differences in the levels of exogenous variables on costs (X^{Qj} versus X^{Qi}), but not in the cost function since costs are evaluated using only the parameters from the efficient cost function, $\hat{C}^{Qi}$.

Any differences in the average cost of assets of the model that cannot be explained by exogenous variables are said to constitute the measure of the inefficiency residual. This is given by the following expression:

$$Ineff = \left[A\hat{C}^{Qj} - A\hat{C}^{Qj*} \right] / A\hat{C}^{Qi} \equiv Diff - Market. \tag{4.4}$$

The thick frontier approach does not impose any assumptions on the distribution of the inefficiencies or the random error. The only assumption is that inefficiencies differ between highest and lowest quartiles, and that random error exists between these quartiles. One major limitation of this method is that it does not provide exact point estimates of efficiency for individual firms, but instead generates an estimate of the general level of overall efficiency.

4.2.3 'Distribution-free' Approach (DFA)

The distribution-free approach is again based on the two-part composed error term (the first part measures efficiency and the second statistical noise) and also specifies a functional form for the frontier. Instead of imposing predetermined distributions on the X-efficiencies and the random error, this approach is based on the notion that efficiency differences across firms are stable over time, and the random error is ephemeral and averages out over time. Good management maximises long-run profits by keeping costs relatively low over long periods of time, although costs may fluctuate because of external shocks and measurement errors.

Under this approach the frontier can be expressed as:

$$C = C\,(Y,\,w,\,v,\,u) \tag{4.5}$$

Following the definition of the frontier, the cost equation for bank *i* is specified as:

$$\ln C_i = \ln C(Y_i, w_i) + \ln v_i + \ln u_i \tag{4.6}$$

where the terms ln v_i and ln u_i are treated as a composite error term.

Within this approach there are two methods to estimate X-efficiency, the average residual method and the bank dummy method [Berger (1993)]. The *average residual method* assumes that the part of the error term that measures inefficiencies is orthogonal to the regressors. The firm with the lowest average cost function residuals is assumed to be fully efficient, hence to lie on the frontier. Estimates of inefficiencies for each bank in a panel data set are defined as the difference between its average residual and the average residual of the firm at the frontier. Following Berger (1993), Bauer *et al.* (1998), and Rogers (1998b), under this approach, given that the random error ln v_i tends to cancel out over time, the X-efficiency measure is given by:

$$x - eff = \exp(\ln \hat{u}_{\min} - \ln \hat{u}_i) \tag{4.7}$$

where $\ln \hat{u}_i$ is the average residuals for the i^{th} bank, and is an estimate of ln u_i; $\ln \hat{u}_{\min}$ is the minimum $\ln \hat{u}_i$, and it assumes that the firm with the lowest average cost function residual is fully efficient.

In cases where the residual accounting for inefficiency is not orthogonal to the repressor, the X-efficiency may be understated. The problem can be resolved using the *dummy variable method*, which drops the orthogonality assumption and is based on a fixed effect model [Berger (1993)]. Thus, instead of using the residuals to estimate efficiency, a firm-specific dummy variable replaces the ln u_i in the cost equation. This dummy variable is constant over time, and takes the values of the efficiency variable plus the average over time of a constant term α. Any differences in the fixed effects estimated across firms are attributed to inefficiencies.

In the distribution free approach inefficiencies can follow almost any distribution, as long as inefficiencies are not negative. But, the major drawback of this method is that changes in efficiency over time brought about by external (such as regulatory reform, interest rates etc.) are considered as average deviations of each firm from the best-average practice frontier, rather than the efficiency at any point in time [see Berger and Humphrey (1997)].

4.2.4 Data Envelopment Analysis (DEA)

The Data Envelopment Analysis is a non-parametric method that uses linear programming techniques to construct a piecewise linear envelope to a set of observed output and input data. This method was named as Data Envelopment Analysis (DEA) by Charnes *et al.* (1978, 1981) although the genesis of the approach lies on the work by Farrel (1957). Farrel considered a single firm that produced one output with the use of two inputs, under the condition of constant returns to scale. The assumption of constant returns to scale allowed a diagrammatic exposition of all relevant information in a simple isoquant. This isoquant represents various combinations of the two inputs that a perfectly efficient firm might use to produce a unit of output. He then extends his analysis to multi-input and multi-output situations. Later studies have extended this approach to the case of variable returns to scale and developed corresponding efficiency measures [Fare *et al.* (1983)].

In their original paper, Charnes *et al.* (1978), introduced the generic term 'Decision Making Units' (DMU) which describes the collection of firms, departments, divisions or administrative units assessed for efficiency which have common input and outputs. According to Charnes and Cooper (1985) a hundred percent efficiency is attained by any DMU only when comparisons with other relevant DMUs do not provide evidence of inefficiency in the use of any input and output.

For the determination of the overall efficiency under the DEA, the following linear programming model is constructed [Aly *et al.* (1990), Farrier and Lovell (1990)]. First, the minimum cost of producing the output of a particular firm is calculated as:

$$\begin{aligned}&\text{Min}\, px\\ \text{subject to}\\ &Y \leq zY\\ &x \geq zX\\ &z \in R^{k}{}_{+}\end{aligned} \tag{4.8}$$

where:

Y = the m dimensional vector of output produced by a particular firm,

x = the n dimensional vector of inputs utilised by a particular firm,

Y = the (k . m) matrix of outputs, where k represents the number of firms,

X = is the (k . m) the matrix of inputs,

z = the vector of the weights or intensity parameters attached to each of the observations or the firms in the determination of the minimum cost,

p = the vector of input prices.

After calculating the minimum cost (MC) the overall efficiency is measured by:

$$OE = MC / C \tag{4.9}$$

where C is the actual cost for a particular firm. Overall efficiency represents the potential or efficient input to actual usage input. This can be decomposed into two components; the technical (T) and the allocative (A) efficiency. Consequently, overall efficiency measures the proportional reduction in costs which could have been obtained if a firm c had been both allocatively and technically efficient. The relationship between overall, technical and allocative efficiency is given by the following:

$$OE = T + A \tag{4.10}$$

A second linear program is constructed to measure the technical efficiency. This is stated as:

$$
\begin{aligned}
&\text{Min } T \\
&\text{subject to} \\
&\quad y \leq zY \\
&\quad Tx \geq zX \\
&\quad z \in R^{k}{}_{+}
\end{aligned}
\tag{4.11}
$$

After the T and OE are calculated, A can be derived through substitution into the equation of the overall efficiency.

The unique feature of the DEA is that a piecewise linear surface rests on top of the observation, instead of trying to fit a regression plane through the centre of the data. The advantage of this method over parametric techniques is that it does not require any assumptions about the functional form. The efficiency of the DMU is measured relative to all other DMUs with the only constraint that all the DMUs, lie on or below the efficient set.

The major drawback of this non-parametric technique is that it generally assumes that there are no random errors, and consequently all deviations from the estimated frontier represents inefficiencies. The presumption of no random error, rests on the assumptions of no measurement error in constructing the frontier and no luck that temporarily gives a decision making unit better measured performance for one period. Finally, it does not take into account inaccuracies generated by *accounting* rules that would make *measured* outputs and inputs diverge from *economic* output and input. Any of these errors that appear in an inefficient unit's data may reflected a change in its measured efficiency [Berger and Humphrey (1997)].

4.3 Literature Review

The efficiency literature on financial institution is both large and recent. As mentioned earlier, empirical evidence suggests that X-inefficiency accounts for 20% or more of the costs in banking, whereas scale and product mix inefficiencies account only for less than 5% of costs. X-inefficiency has also been found to have a strong empirical association with higher probabilities of financial institution failures over several years following the observation of substantial inefficiency. In general, the empirical literature on the X-efficiency has focused mostly on the US banking industry, and only a handful of studies have examined other regions.

4.3.1 Comparison of different efficiency measurement methods

Despite intense research efforts there is no consensus on the best method or set of methods for measuring frontier efficiency. Most studies have applied a single efficiency approach, while only few studies have compared two or more methods by use of the same data set.

Studies that concentrate on a single parametric method suggest that the efficiency results appear to be consistent over the years [Berger and Humphrey (1992)]. Also, these efficiencies are found to be related in the expected way with standard non-frontier measures of performance, such as the return on assets [Berger and Mester (1997), Bauer *et al.* (1998)].

Studies that compare two parametric methods indicate that both methods produce similar results for overall efficiencies. Bauer *et al.* (1993) employ both the SFA and the TFA on a US data sample. Their results indicate a 15% average inefficiency for the sample period by the application of SFA. The second method, TFA, produces similar results with the level of inefficiencies amounting to 18%. They conclude that, although the two approaches rank individual banks quite differently, they yield similar average efficiency findings. Berger and Mester (1997) utilise DFA and SFA and suggest that the choice of measurement method makes very little difference in terms of either average industry efficiency or ranking of individual firms. They find a rank order correlation of 0.988 between the two techniques.

Proceeding with the comparison of parametric and non-parametric studies, the findings suggest fairly close efficiency estimates. Ferrier and Lovell (1990) employ the DEA and SFA methods. The results form the DEA indicate an overall average inefficiency of 21%, while the SFA results suggest an overall inefficiency of 26%. However, they find a rank order correlation of only 0.02 between the two methods. On the other hand, Resti (1997) find a very high rank order correlation between the DEA and the SFA of 0.73 to 0.83.

Bauer *et al.* (1998) apply all four approaches to a single data set. The findings yield some mixed evidence. The SFA, TFA, and DFA produce consistent results, i.e. the same distributions of efficiency, rank banks in approximately the same order, and identify mostly the same banks as best- and worst-practice. On the other hand, the DEA yields much lower average efficiencies, while the ranking of banks and the identification of best and worst banks are different from parametric methods. A common feature is that all methods are consistent over time.

4.3.2 Deregulation and efficiency

With the major regulatory changes that occurred during the last two decades, regulators aimed to improve efficiency of the banking institutions; however the empirical literature on the matter produces mixed results.

A portion of US studies suggest that the banking efficiency remains relatively unchanged to deregulation [Bauer, Berger, and Humphrey (1993)], while others suggest the opposite. Grabowski *et al.* (1994) claim that deregulation forced inefficient banks to leave the industry, while overall average results reflect little change in efficiency. Elyasiani and Mehdian (1995) indicate that small and large banks are affected differently by changing market conditions. Based on group specific frontiers the results show that in the pre-regulation environment small banks have been more efficient than large banks, while after deregulation both groups become equally efficient. Berger and Mester (1997) find evidence of some decrease in cost efficiency between 1980s and 1990s, while large banks show a considerable decline in profit efficiency.

Contrary to the aforementioned findings, Berg *et al.* (1992) state that Norwegian banks experienced improved efficiency and productivity after deregulation. This is also the case with the Turkish banking industry [Zaim (1995)], which has benefited from the more liberalised banking environment.

In conclusion, deregulation does not always improve efficiency. Industry conditions and market structure prior to deregulation are important and might produce different results. Nevertheless, measurement over longer time periods may eventually show a net improvement in efficiency. As Berger and Mester (1997) noted, competitive pressures on banks result in increasing output quality, which in turn raises banks' costs, while expecting long-run cost and revenue benefits from higher quality. Additionally, new technology has changed the way banks operate and the fixed costs of this new technology have not yet been recovered.

4.3.3 Problem loans, risk and efficiency

The primary objective of regulations is to ensure soundness the banking industry. '*Most bank failures are directly related to having a large number of problem loans, a low capital position, weak or negative cash flow, and a poor management quality*'[55]. Bank studies that have examined these issues find evidence that institutions display low efficiency prior to failure, together with positive relationship between management quality and efficiency.

A number of studies find evidence that failing banks tend to be located far from the best frontier, indicating that they perform at low cost efficiency [Berger and Humphrey (1992), DeYoung and Whalen (1994)]. Berger and DeYoung (1997) utilise Granger-causality techniques and find that problem loans precede reductions in measured cost efficiency. Also, their results indicate that reductions in capital at thinly capitalised banks precede increases in problem loans.

Some efficiency studies include measures of problem loans and risk. Mester (1996) include the average volume of nonperforming loans in the cost function to capture the quality of banks, and the average volume of equity capital per bank as a measure of bank risk. Her results indicate that the inclusion of these variables lad to an increase in efficiency.

4.3.4 Mergers and acquisitions and efficiency

One of the major structural changes that have occurred in the banking and financial services industries is the significant reduction in the number of independent financial institutions. This has partly been the result of the increased number of mergers and acquisitions (M&A). One of the motives for mergers is cost efficiency improvements. Cost efficiency could be considerably improved by a merger where a relatively efficient bank acquires a relatively inefficient bank and spreads its superior management talent over more resources.[56]

The first studies on the effect of M&A activities on the performance of banks employed financial ratio analysis. Most of these studies suggest that there are no potential benefits on costs from merger activities [Rhoades (1986, 1990), Cornett and Tehranian (1992), Linder and Grane (1992), Srinivasan and Wall (1992)]. However, these studies faced important methodological problems. The utilisation of simple cost ratios to account for efficiency does not take into account differences in input prices and output mix.

In an attempt to overcome the aforementioned limitation, later studies employ more sophisticated econometric techniques. Nevertheless, cost X-efficiency studies that utilised 1980s data produced the same results as the cost ratio studies. Their findings indicate no or very little cost improvement from M&As during the 1980s [Berger and Humphrey (1992), Rhoades (1993), DeYoung (1997)]. Studies using early 1990s data generate mix results. Rhoades (1998b) finds modest efficiency gains, while Berger (1998) finds very little improvement in average cost efficiency for M&As of both large and small banks. A study on the European banking markets by Molyneux *et al.*

[55] Berger and Humphrey (1997), p.p.23.
[56] See Berger, Hunter and Timme (1993), and Akhavein *et al.* (1997) for more discussions of this issue.

(1996), found evidence that M&As between large banks, either within domestic banking markets or across national borders can create substantial cost savings or cost reductions depending on the merger partners chosen.

The most recent of studies utilise the profit function approach to capture the effects of M&As on performance [Akhavein *et al.* (1997), Berger (1998)]. These studies find that M&As actually improve profit efficiency. An advantage of the profit efficiency over cost efficiency is that it embodies the scale, scope, product mix as well as X-efficiency effects, which maybe the reason for the different results.

4.3.5 A summary of the evidence

Estimates of efficiency of financial institutions vary substantially depending on the data source and on the efficiency concepts and measurement methods used in the studies. However, once a frontier approach is adopted and an output specification is selected, efficiency estimates are fairly stable over time, showing a level of persistency [Berger and Humphrey (1997)].

In general, the estimates from parametric studies generate similar estimates to those that apply non-parametric methods, although the later yield slightly lower mean efficiency estimates. Berger and Mester (1997b) make an attempt to determine the importance of different efficiency concepts and the measurement techniques applied to the findings of the studies. They apply multiple efficiency concepts and use a number of different measurement methods. From their results it is generally clear that the choice made concerning efficiency measurement makes little difference to empirical findings. This, in turn, suggests that efficiency estimates are robust to different methodologies.

Table 4-1 provides a summary of the studies of bank frontier in chronological order.

Table 4-1: Summary of Bank Frontier Studies.

Authors	Methodology & Data Set	Conclusions
Aly, Grabowski, Paruska, and Ragan (1990)	DEA 322 US banks 1986	Low overall efficiency, 65%. More allocative than technical efficient. Efficiency differences exist between branching and non-branching banks.
Ferrier and Lovell (1990)	DEA & SFA 575 US banks - 1984	Overall Inefficiency: DEA 21%, SFA 26%. More allocative than technical efficient.
Berger and Humphrey (1991)	TFA 13,951 US banks 1984	Overall inefficiencies: Branch 24.8%, Unit 20.2%. Inefficiencies greater for small banks. Technical inefficiencies dominate allocative.
Berger and Humphrey (1992)	TFA 14,000 US banks 1980, 1984, 1988	Overall Inefficiencies: Branch 9.9%, 12.7%, 22.6%. Unit 14.6%, 14%, 21.3%. Increase of inefficiencies over the period due to non adjustment to deregulation.
Bauer, Berger, and Humphrey (1993)	SFA & TFA 683 US banks 1977 to 1988	Average inefficiencies: SFA 18%, TFA 21%. Consistent estimates between the two approaches.
Grabowski, Rangan, Rezvanian (1993)	DEA 4,094 US banks 1989	Overall efficiency 68%, Allocative 94%, Technical 72%.
English, Grasskopf, Hayes, and Yaisawarng (1993)	DEA 442 US banks 1982	Banks in the sample both technical and allocative inefficient; hence, not revenue maximisers.
Faver and Papi (1993)	DEA 174 Italian banks 1991	Higher efficiency of banks engaged in non-traditional activities. Overall efficiency, Intemediation approach 87.8%, Asset approach 79.4%
Berg, Forsund, Suminen, Hjalmarsson (1993)	DEA 503 Finnish, 126 Swedish, 150 Norwegian - 1990	Swedish banks are the most efficient and only one Finish and no Norwegian bank have efficiency score above 90%.
Mester (1993)	SFA US S&Ls 1991	Transfer from mutual to stock S&Ls suggests that deregulation of interest rates and increased competition affected mutual S&Ls. On average stock S&L are less efficient than mutual S&L.
Berger, Hancock,	DFA	Overall efficiencies:

and Humphrey (1993)	US banks 1984 to 1989	Unit branching 52%, Limited branching 65%, Statewide branching 66%. Technical inefficiencies dominate allocative. Output inefficiencies greater than input inefficiencies.
Berger (1993)	DFA 900-1,010 US banks 1980-1989	Overall inefficiencies: 46% for Unit, 46% for Limit, and 32% for State banks. Due to the variations of estimates from year to year, there are doubts about the ability of any single period method in obtaining accurate results.
Kaparakis, Miller, and Noulas (1994)	SFA 5,548 US banks 1986	Banks become less efficient as size increases. Further, the more branches a bank operates the lower the cost efficiency. Overall efficiency 90%.
Grabowski, Rangan, and Rezvanian (1994)	DEA 669 US banks 1979, 1983, and 1987	Banks are more allocative than technically efficient – pure technical being the source of inefficiencies. Overall inefficiencies: *1979*: 74%, *1983:* 76%, *1987:* 73%. The results do not support the hypothesis that deregulation favour the efficiency of banks.
Drake and Howcroft (1995)	DEA 190 UK clearing bank branches	Only 107 out 190 are found to be efficient, while overall efficiency amounted to 92%. The results suggest considerable diversity across branches in terms of efficiency levels.
Elyasiani, and Mehdian, (1995)	DEA 150 small and 150 large US banks 1979 and 1986	Efficiency of small and large banks is quite similar. But they are affected differently on average and small banks are affected differently within the group. The average small bank has to 'struggle' to keep up with the changing market conditions. Overall efficiency: Small B. 95%, 92%, Large B. 96%, 97%.
Miller, and Noulas (1996)	DEA 201 US banks 1984 to 1990	Overall inefficiency 5%. This is due to the nature of the sample, i.e. very large and profitable banks which are not affected by changes in the market structure.
Mester (1996)	SFA 214 US banks	Includes two extra variables in the cost function.

	1991 to 1992	The first one is the average volume of non-performing loans, to capture the quality of the banks, and the second is the average volume of equity capital per bank, as a measure of bank risk. Overall inefficiency 6-9%.
Berger, and Mester, (1997a)	SFA & DFA 6,000 US banks - 1990 to 1995	The utilisation of fourier specification, does not generate different results from the translog. DFA: cost function: fourier-86.8%, translog-86% efficiencies; Standard profit function: 54.9%, 53.9%, Alternative profit function: 46.3%, 45.2%, SFA: cost function 94.2%, alternative profit function 53.1%.
Berger, and DeYoung (1997)	SFA US banks -1985 to 1994	Overall efficiency 92%. The applied Fourier-flexible, truncated normal specification applied, which is more general than the standard translog, half-normal model typical employed in the literature. This according to the authors generates betters estimates.
Berger and Mester (1997b)	DFA All US commercial banks 1984-1995	Cost efficiencies for all sizes average about 80% over *1984-1989,* and decline to 77% over *1990-1995.* Profit efficiency for the first sub-period is the same for all bank sizes, while the second period large banks are considerably less efficient than small banks.
Akhaven, Berger, and Humphrey (1997)	DFA US banks 1980-1990	Bank mergers in the 1980s significantly improve profit efficiency. The average profit efficiency rank of merging banks increased from 74th percentile to the 90th percentile of the group of large banks.
Berger, Leusner, and Mingo (1997)	DFA 832 branches of a single US bank 1989-1991	Most branches are smaller than efficient scale, and there are roughly twice as may branches as needed to minimise bank costs. X-inefficiencies are about 5% to 10% of total branching costs, and 20%-25% of branch operating costs.
Athanassopoulos, Soteriou, and	DEA UK, Greek and Cypriot	UK branch network exhibits an overall dominance over the other

Zenios (1997)	single bank branches	networks. Cypriot network is less efficient than the Greek one. Greek and Cypriot network contain mechanisms that can be used by UK branches to improve further their efficiency.
Bauer, Berger, Ferrier, Humphrey (1998)	SFA, TFA, DFA, & DEA 683 US banks 1977-1988	They examine the consistency of the estimates produced from the 4 methods. The parametric techniques produced consistent results, as opposed to the DEA A common feature, is that all methods were consistent over time.
Rogers (1998b)	DFA More than 10,000 US banks 1991-1995	He estimates a translog cost and revenue, as well as profit frontier, each including non-interest income as a measure of non-traditional output. Cost and profit efficiency increase once traditional activities are accounted for, suggesting that the sale of non-traditional activities increase banks' efficiency.

Notes: SFA: Stochastic econometric frontier approach.
TFA: Thick frontier approach.
DFA: Distribution-free approach.
DEA: Data envelopment analysis.

4.4 Methodology.

To measure the X-efficiency of the UK banking institutions, the stochastic econometric frontier model will be applied here. The major advantage of the SFA over the other frontier methods is that it *always ranks* the efficiencies of firms in the same order as their cost function residuals, regardless of the distributional assumptions imposed. Accordingly, banks with lower costs for a given set of input prices, and any other cost regressors will always be ranked as more efficient. This is because the conditional mean or mode of u (inefficiencies, given that the error is given by $\varepsilon_i = v + u$ is always increasing with the residuals [Bauer *et al.* (1998)]. This renders the SFA appealing for regulatory purposes since a firm is measured as high in the efficiency rankings if it keeps costs relatively low for its given exogenous conditions.

Considering the other three approaches, the thick frontier approach is rejected because it does not provide exact point estimates of efficiency for individual firms, but only generates an estimate of the general level of overall efficiency. Additionally, it is based on a rather arbitrary assumption; namely that the lowest average cost quartile within each size class is an adequate thick frontier of efficient firms. The distribution-free approach is based on the notion that inefficiencies of firms are stable over time. Again, this assumption seems unrealistic in banking. Furthermore, it considers changes in efficiency over time due to outside factors (such as regulatory reform, interest rates etc.) as average deviations of each firm from the best-average practice frontier, rather than the efficiency at any point in time [see Berger and Humphrey (1997)]. Finally, the data envelopment analysis does not take into account random error. However, we cannot assume that all the deviations from the estimated frontier represent inefficiencies.

The stochastic frontier is chosen as the best method of estimation, although it is based on the strict assumption that inefficiencies are half normally distributed. To test the effect of this assumption on the robustness of the results we are going to estimate three models that are based on different distributional assumptions. The first model is based on a half-normally distributed error term, a second model is based on the truncated specification, and a third one on the exponential distribution as a special case of the gamma distribution.[57]

Furthermore, according to Berger and Mester (1997), choices concerning efficiency measurement usually make little difference to empirical findings in terms of either average industry efficiencies or rankings of individual firms, suggesting that the efficiency estimates are fairly robust to differences in the methodology. These authors argue that there is no best frontier method, since the true level of efficiency is unknown, and suggest a possible solution to add more flexibility to parametric approaches and introduce a degree of random error into the non-parametric approach.

Now, according to the stochastic econometric frontier model a firm's observed cost will deviate from the cost frontier because of random noise and possible inefficiencies. This approach modifies a standard cost function to allow inefficiencies to be included in the error term. The model consists of a two-part composed error term, where one part accounts for inefficiencies and the other one for statistical noise. This model is based on specific distributional assumptions to separate the two components of the error term. Because inefficiencies can increase cost above frontier levels, they are assumed to be drawn from a one-sided distribution, usually the half-normal. On the other hand, random fluctuations can both increase or decrease costs, hence they are assumed to be drawn from a symmetric distribution, usually the normal.

The methodology for estimating a stochastic or efficient frontier can be outlined as follows. We first considered a stochastic cost function model:

$$TC_i = TC(Y_i,\ w_i) + \varepsilon_i \tag{4.12}$$

where:

TC_i = the observed total cost of firm i,

Y_i = a vector of outputs,

w_i = an input price vector, and

[57] The half normal distribution is a special case of the truncated normal in which the corresponding normal distribution is truncated at the centre.

ε_i = the error term.

Following Ainger, Lovell and Scmidt (1977) we assume that the error term of a cost function can be decomposed into:

$$\varepsilon_i = v_i + u_i \tag{4.13}$$

where:

v_i = random error, and

u_i = inefficiencies.

The error component v_i represents the symmetric disturbance, which is assumed to be independently and identically normally distributed with zero mean and σ_v^2 variance, that is $v_i \sim N(0,\sigma_v^2)$. The error component u_i is assumed to be distributed independently of v_i and to satisfy $u_i \leq 0$, that is $u_i \sim |N(0,\sigma_u^2)|$. Given the distributional assumptions for the error terms v_i and u_i, the density function of the composite error term, $\varphi(\varepsilon_i)$ can be written as [Aigner *et al.* (1977)]:

$$\varphi(\varepsilon_i) = \frac{2}{\sigma} f\left(\frac{\varepsilon_i}{\sigma}\right)[1 - \Phi(\varepsilon_i \lambda \sigma^{-1})] \qquad -\infty \leq \varepsilon_i \leq +\infty \tag{4.14}$$

where:

$\sigma^2 = \sigma_u^2 + \sigma_v^2$,

$\lambda = \sigma_u / \sigma_v$, and

$\varphi(.)$ and $\Phi(.)$ = are the standard normal density and distribution function respectively.

This density is asymmetric around zero, with mean and variance given by:

$$\begin{aligned} E(\varepsilon_i) &= E(u_i) = -\frac{\sqrt{2}}{\sqrt{\pi}}\sigma_u \\ V(\varepsilon_i) &= V(u_i) + V(v_i) = \left(\frac{\pi - 2}{\pi}\right)\sigma_u^2 + \sigma_v^2 \end{aligned} \tag{4.15}$$

The relevant log-likelihood function for a sample N observation is:

$$\ln L = -(N/2)(\ln 2\pi + \ln \sigma^2) + \sum_{i=1}^{N}\left[\ln \Phi(-\varepsilon_i \lambda / \sigma) - \frac{1}{2}(\varepsilon_i \lambda / \sigma)^2\right] \tag{4.16}$$

Jordow *et al.* (1982) have shown that the standard deviation, σ, for u_i and v_i can be used to measure a firm's relative inefficiency, where $\lambda = \sigma_u / \sigma_v$ is a measure of variation originated from inefficiency relative to noise from the sample. Estimates of the model can be computed by maximising the likelihood function directly. They further suggest that observation specific estimates of inefficiency, u_i, can be calculated by using the distribution of the efficiency term conditional on the estimate of the composed error term. The mean of this conditional distribution for the half-normal model is:

$$E(u_i / \varepsilon_i) = \frac{\sigma\lambda}{1+\lambda^2}\left[\frac{\phi(\varepsilon_i \lambda / \sigma)}{1-\Phi(\varepsilon_i \lambda / \sigma)} + \left(\frac{\varepsilon_i \lambda}{\sigma}\right)\right] \tag{4.17}$$

In accordance with Greene (1990) the half-normal distribution proposed by Aigner *et al.* (1977), is somewhat inflexible as it is a single parameter distribution and assumes that the density function of the disturbances is most concentrated around zero. In view of this, Stevenson (1980) suggests shifting the half-normal distribution by allowing a nonzero mode, producing a general truncated normal distribution instead. He analyses the case in which u, still truncated at zero, has a truncated normal distribution with parameter μ (allowed to differ from zero in either direction) and σ_u^2. The conditional distribution for the truncated normal model can be obtained by replacing $\varepsilon_i \lambda / \sigma$ in equation (4.17) with the following expression:

$$\mu^* = \frac{\sigma_v^2 \mu + \sigma_u^2 \varepsilon_i}{\sigma_v^2 + \sigma_u^2} . \left[\frac{\sigma_v^2 . \sigma_u^2}{\sigma_v^2 + \sigma_u^2}\right]^{-1/2} = \frac{\varepsilon_i \lambda}{\sigma} + \frac{\mu}{\sigma\lambda} \tag{4.18}$$

where μ is the mean of the non-truncated distribution. Its log-likelihood function is given by:

$$\ln L = \ln\sigma - \frac{1}{2}\ln\left(\frac{2}{\pi}\right) - \frac{1}{2}\left(\frac{\varepsilon_i - \mu}{\sigma}\right)^2 + \ln\Phi\left(-\frac{1}{\sigma}\right)\left(\varepsilon_i\lambda + \frac{\mu}{\lambda}\right) - \ln\Phi\left(\frac{(-\mu/\sigma)(1+\lambda^2)^{1/2}}{\lambda}\right) \tag{4.19}$$

Finally, we also computed the estimates for the exponential model, which emerges when P is constrained to be equal to one in the gamma model[58]. The mean of the conditional distribution is as follows [Greene (1990)]:

$$E(u_i \mid \varepsilon_i) = z + \sigma_v\left(\frac{\phi(z/\sigma_v)}{\Phi(z/\sigma_v)}\right) \tag{4.20}$$

where: $z = \varepsilon - \theta\sigma_v^2$.

The log-likelihood function of the exponential model is equal to:

$$\ln L_i = \ln\theta + \frac{1}{2}\theta^2\sigma_v^2 + \theta\varepsilon_i + \ln\Phi\left(\frac{-\varepsilon_i}{\sigma_v} + \theta\sigma_v\right) \tag{4.21}$$

Going back to the cost function, the single equation stochastic cost function model for bank i at time t is specified as:

$$\ln TC_{it} = \ln C_t(Y_{it}, w_{it}) + \ln\varepsilon_{it} \tag{4.22}$$

where:

TC denotes total cost (operating and interest cost) for the bank i, and

C(Y, w) is a cost function with output quantity vector Y and input price vector w.

In this chapter, a translog cost model is going to be applied to measure the efficiency of the UK banking institutions. As mentioned in the previous chapter, the translog specification overcomes the main drawbacks of the other major cost function. Accordingly, when the minimum total cost function (the cost frontier) is translog, then a bank's observed total cost is the following:

$$\ln TC = \alpha_o + \sum_{i=1}^{n} \alpha_i \ln w_i + \sum_{k=1}^{m} \beta_k Y_k + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} \ln w_i \ln w_j +$$
$$\frac{1}{2}\sum_{k=1}^{m}\sum_{\zeta=1}^{m} \delta_{k\zeta} \ln Y_k \ln Y_\zeta + \sum_{i=1}^{n}\sum_{k=1}^{m} z_{ik} \ln w_i \ln Y_k + u_i + v_i \qquad (4.23)$$

where:

w_i, w_j, = price of input i and j,

Y_k Y_ζ = outputs k and ζ,

TC = total cost,

v_i = random error, where $v_i \sim N(0,\sigma_v^2)$,

u_i = inefficiencies, where $u_i \sim |N(0,\sigma_u^2)|$, and

α_i, β_k, γ_{ij}, δ_{kj}, z_{ik} = are independent parameters.

Next, equation (4.23) is estimated, by use with three models estimated for each year of the period 1981-1995. *Model I* is based on the assumption of half normally distributed u_i, *Model II* on the truncated distributional assumption, and *Model III* on the exponential distribution. The three different distributional assumptions are utilised to test the robustness of the estimates to different specifications.

[58] For details about the Gamma stochastic frontier model see Greene (1990).

4.5 Data sample and determination of bank input and bank output

As be seen in the previous chapter, the different determination of input and output does not yield great variation in the results. As shown in Chapter 3, the inclusion of security investments increases the R^2 of the model for most years. Hence, in the current chapter we employ the value-added approach for the determination of input and output. The sample is the same as in Chapter 3.

Total costs (TC) are proxied as the sum of all operating cost incurred by the bank in the production of output and services together with interest payable. Consistent with the value-added approach all outputs are measured in pounds sterling.

The inputs selected are: labour (x_1) and physical capital (x_2). Labour is measured by the total number of employees at the end of the time period, while physical capital by the net book value of total fixed assets.

The price of labour (w_1) is derived by taking the total expenditure on employees divided it by the number of employees. A proxy for the price of capital (w_2) is calculated by dividing the total expenditure on fixed assets with the net book value of fixed assets. Three outputs are used in the current study, all measured in sterling pounds, Total Deposits (Y_1), Total Loans (Y_2), and Security Investments (Y_3).

4.6 Empirical Results.

As already mentioned, the cost function is estimated three on the basis of three alternative distributional assumptions about the part of the error term that captures the inefficiencies. *Model I* is a normal half- normal model, where random disturbances follow a normal distribution and inefficiencies, consistent with the stochastic frontier approach, follow a half-normal distribution. *Model II*, is a more flexible model, where the random error is normally distributed, but inefficiencies are truncated normal. *Model III* is the exponential model that follows an exponential distribution. A maximum likelihood procedure was used for the derivation of the estimates. Estimated parameters for all models are shown in the Appendix for Chapter 4 (Tables 4-6 to 4-8).

Estimating the stochastic cost frontier (4.24) we obtaine a set of valuesfor u_i, (inefficiency), for each firm, each year. Table 4-2 presents the average inefficiency values, for the years 1981 to 1995 after estimating *Model I*, Table 4-3 after estimating *Model II*, and Table 4-4 after estimating *Model III*.

Table 4-2: Inefficiency values obtained from *Model I* – Half-normal distribution.

Year	Mean	Median	Standard Deviation	Minimum	Maximum
1995	0.217	0.188	0.098	0.058	0.451
1994	0.212	0.197	0.073	0.086	0.472
1993	0.260	0.245	0.124	0.033	0.619
1992	0.267	0.235	0.130	0.080	0.584
1991	0.283	0.253	0.151	0.045	0.763
1990	0.228	0.213	0.081	0.100	0.461
1989	0.173	0.169	0.054	0.091	0.385
1988	0.171	0.159	0.061	0.067	0.431
1987	0.272	0.193	0.263	0.006	1.010
1986	0.270	0.208	0.244	0.005	0.984
1985	0.285	0.200	0.252	0.002	0.867
1984	0.258	0.216	0.168	0.056	0.994
1983	0.279	0.242	0.262	0.004	0.883
1982	0.198	0.178	0.090	0.071	0.615
1981	0.215	0.191	0.103	0.062	0.707

Table 4-3: Inefficiency values obtained from *Model II* –Truncated distribution.

Year	Mean	Median	Standard Deviation	Minimum	Maximum
1995	0.203	0.192	0.068	0.084	0.350
1994	0.212	0.197	0.073	0.086	0.472
1993	0.255	0.235	0.125	0.074	0.633
1992	0.213	0.205	0.100	0.028	0.448
1991	0.272	0.241	0.135	0.049	0.624
1990	0.228	0.213	0.081	0.100	0.461
1989	0.114	0.111	0.032	0.066	0.253
1988	0.184	0.161	0.077	0.066	0.514
1987	0.284	0.172	0.245	0.041	1.354
1986	0.287	0.217	0.179	0.044	0.761
1985	0.235	0.176	0.151	0.076	0.756
1984	0.258	0.216	0.168	0.056	0.994
1983	0.288	0.196	0.296	0.004	1.386
1982	0.183	0.176	0.094	0.045	0.443
1981	0.213	0.189	0.101	0.061	0.694

Table 4-4: Inefficiency values obtained from *Model III* – Exponential distribution.

Year	Mean	Median	Standard Deviation	Minimum	Maximum
1995	0.196	0.190	0.079	0.0519	0.423
1994	0.188	0.132	0.145	0.0717	0.816
1993	0.186	0.156	0.108	0.0625	0.727
1992	0.218	0.174	0.149	0.0766	0.984
1991	0.260	0.226	0.139	0.0741	0.633
1990	0.223	0.177	0.139	0.0770	0.719
1989	0.104	0.098	0.037	0.0562	0.289
1988	0.179	0.156	0.069	0.0657	0.470
1987	0.279	0.166	0.314	0.0001	1.496
1986	0.297	0.197	0.224	0.0010	0.979
1985	0.267	0.181	0.249	0.0012	0.871
1984	0.267	0.241	0.174	0.0435	1.099
1983	0.281	0.159	0.308	0.0012	1.407
1982	0.190	0.171	0.113	0.0596	0.727
1981	0.202	0.177	0.126	0.0468	0.759

The level of mean inefficiencies for each frontier model over our 15-year period (1981-1995) is displayed in Figure 4-1.

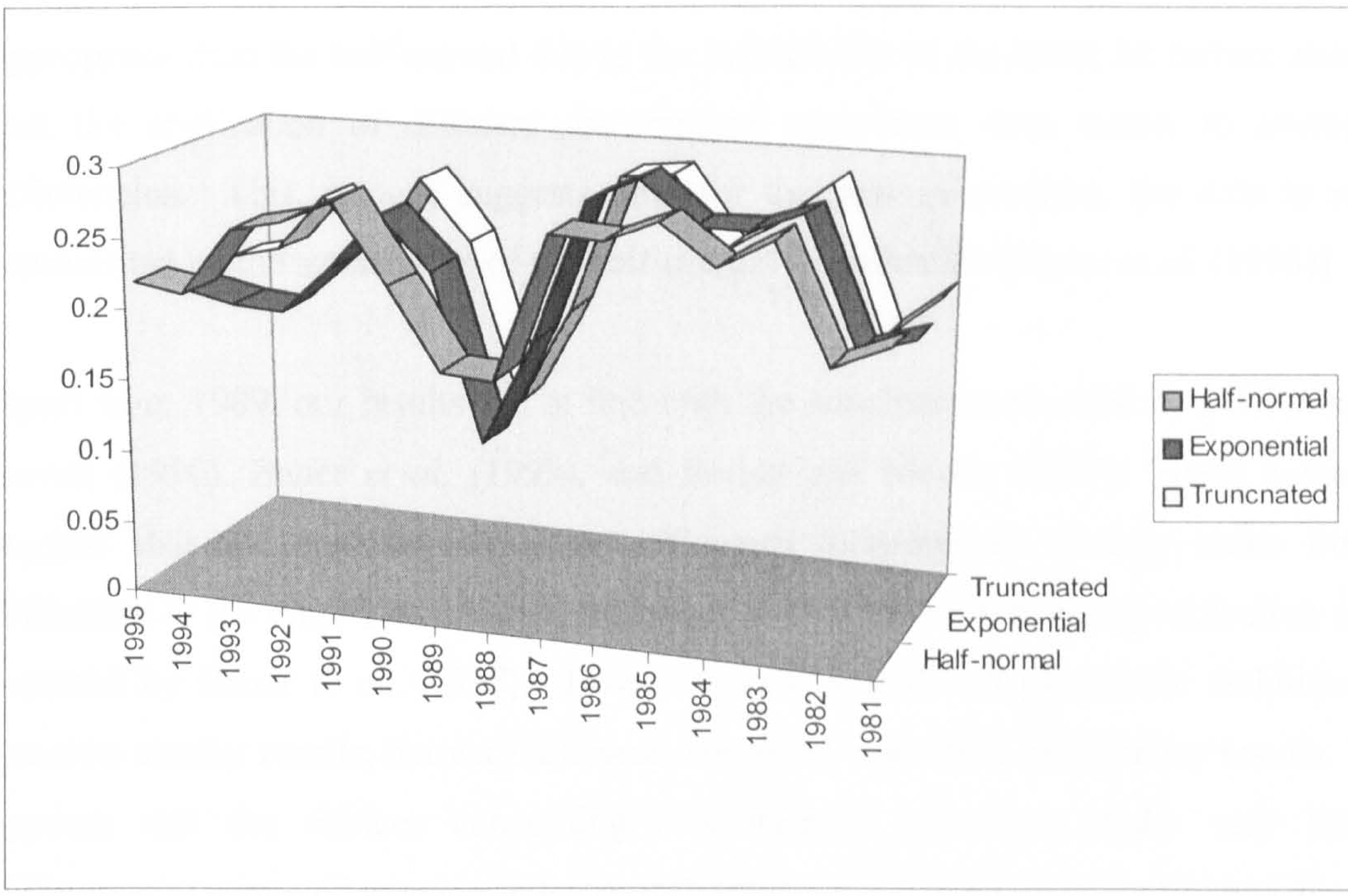

Figure 4-1: Average inefficiencies over time for the three techniques.

Although the literature on the efficiency of financial institutions has been quite extensive during the past few years, only few studies have compared multiple techniques to examine the sensitive of the estimates based on different efficiency methods [see Ferrier and Lovell (1990), Bauer *et al.* (1998), Berger and Mester (1997), Bauer *et al.* (1998)]. Moreover, all these studies applied the SFA by utilising the standard normal/half-normal assumption. The current study proceeds a step further and examines the robustness of the results by use of different distributional assumptions.

Results reported in Tables 4-2 to 4-4 together with Figure 4-1, indicate small differences in average efficiencies and in efficiency dispersion when alternative distributional specification are used. The only exception is for the year 1989, where the truncated and exponential models report an inefficiency level of 11% and 10% respectively, whereas the half-normal model gives an inefficiency level of 17%. An explanation may be the inflexible specification of the half-normal model, as it embodies the assumption that the density function of the disturbances is concentrated near zero, whereas the other two models are considered to be more flexible. As Greene (1990) points out, alternative distributions for efficiency may be more appropriate than the half-normal due to the inflexibility of the latter; he further stated that the application of different distributions sometimes does matter to average efficiencies. This, in turn suggests, that for the year in question, the data is not represented by the assumption of the half-normal specification [Bauer *et al.* (1998)].

Apart from 1989, our results are in line with the conclusions reached by Ferrier and Lovell (1990), Bauer *et al.* (1993), and Berger and Mester (1997). These authors suggest that the choices concerning efficiency measurement usually make little difference to the empirical findings. Substantial differences between specification are reported by Bauer *et al.* (1998). However, in this study all parametric techniques generate similar results, but very different compared with DEA inefficiency results. It appears that the choices concerning measurement techniques make very little difference in terms of average industry efficiency or the rankings of individual firms in the current data set. Thus, we conclude that the efficiency estimates are fairly robust to differences in methodology.

Following the Bauer *et al.* (1998) we proceed with the correlation of the different inefficiency measures obtained from different distributional specifications. These are reported in Table 4-5.

Table 4-5: Correlation among the inefficiency scores.

Distribution	Half-normal	Exponential	Truncated
Half-normal	1		
Exponential	0.838	1	
Truncated	0.859	0.910	1

Table 4-5, in accordance with the aforementioned stated, suggests that there is a high correlation between the results obtained from the three models. The highest correlation (0.91) is between the exponential and truncated models.

Now, going back to Tables 4-2 to 4-4, and Figure 4-1, starting from 1981 and 1982 the average level of inefficiency is around 20%, and rise rapidly the next year to an average of 8% and remains fairly constant at around 28% until 1987. This was in line with the national and international economic events. In 1983 a notable number of countries experienced debt-servicing difficulties. This resulted in an increase in bad debt provision accounts to banks engaged in business with these countries. An extra increase in provision came into effect in 1987. During that period, there was a modest growth of world trade together with a difficult economic situation in several countries (including the UK) that experienced a moderate rate of growth. Due to the worsening economic conditions, banks' loans remain constant over this period (see Chapter 1). Given that loans constitute a major source of income for the banking industry, this stable pattern together with the increases in bad debt provisions, lowers banks' profits, and this in turn raises costs. Furthermore, the global changes in the financial markets together with the internalisation of financial services and deregulation pattern raise the competitive pressures in the industry. This leads banks to restructure their balance sheet and to undertake heavy investment in information technology in conjunction with higher depreciation charges, and to increase output quality. All these introduce temporarily increased costs to the banking industry, resulting in the high levels of inefficiency.

In 1988 the inefficiency level started to drop, reaching its minimum level in 1989. An explanation of this result maybe found in the considerable increase in residential and personal loans, which in turn increases the level of output. Nevertheless, in the next decade the inefficiency levels start to rise representing the difficult trading conditions for that period. Moreover, the beginning of the new decade is characterised by increased competition and in the light of the second banking directive and the single economic market several banks are reviewing their activities. The UK economy starts to recover only after 1993 where this was again reflected in the inefficiency estimates. By 1995 the inefficiency level for the UK industry amounts to 20%. Considering an average for the whole period, the inefficiency levels range between 18 to 28 % with an exception in 1989.

The heavy costs that banks experienced during the 1980s start to diminish with the beginning of the next decade, as inefficiency levels are around 20%. In line with the results of Berg *et al.* (1992), Zaim (1995), and Rogers (1998b) our results suggest that in the 1990s UK banks started to realise the benefits from deregulation, where this deregulation helped banks to reduce their inefficiencies level. On the other hand Berger and Mester (1997, 1999) suggest that there is a slight increase in inefficiency from 1980s to 1990s in the US banking industry. Nevertheless, the US banks, which are more heavily regulated, might have not fully adjusted to the changes in the business conditions brought about by deregulation. In the UK, always in the author's best knowledge, efficiency on bank level has never been tested; only two studies at branch level suggest an 85 – 90% of efficiency [Athanassopoulos (1995, 1997)].

4.7 Conclusions

The literature on financial institution efficiency is both large and recent. On average X-inefficiency accounts for a considerable portion of cost, and is a greater source of performance problems than scale and product mix inefficiencies, while a number of studies find evidence that failing banks perform at high cost inefficiencies. Moreover, frontier efficiency is superior for most regulatory and other purposes than standard ratio analysis commonly used by regulators, financial analysts, and industry consultants when trying to assess performance, as it permits overall objectively determined numerical efficiency values and ranking of firms.

In the current study the efficiency of the UK banks for the period 1981 to 1995 has been examined by estimating an efficient frontier and by measuring the average differences between observed banks and banks at the frontier. Three different model specifications and applied to check the robustness of estimates to different specifications.

Empirical findings are in line with those of the current literature and indicate that X-inefficiencies account for 20 to 30 % of costs in UK banking. Moreover, the results, consistent with other studies, suggest that choices made concerning measurement techniques make very little difference in terms of average industry efficiency or rankings of individual firms in the current data set. Thus, the efficiency estimates are fairly robust to differences in methodology and are in line with those obtained from US banking studies. Finally, in line with the results of Berg *et al.* (1992), Zaim (1995), and Rogers (1998b) our results suggests that in the 1990s UK banks started to realise the benefits from deregulation, which helped banks to reduce their inefficiencies level.

4.8 Appendix to Chapter 4

Table 4-6: Parameter estimates from *Model I* - Half-normal distribution.

Variable	1995	1994	1993	1992	1991	1990	1989	1988
Constant	4.289	3.891	-7.081	-2.051	5.251	8.943	10.444	1.658
	(9.86)	(24.75)	(5.43)	(5.08)	(9.29)	(4.99)	(9.84)	(7.09)
w_1	0.256	-0.920	0.844	1.560	-2.748	-1.578	-1.417	0.636
	(2.9)	(3.39)	(1.36)	(1.381)	(2.44)	(1.39)	(3.01)	(1.97)
w_2	-0.2E-02	0.453	0.203	-0.531	-1.156	0.063	-0.240	-1.262
	(1.27)	(1.71)	(1.20)	(1.00)	(1.25)	(0.74)	(0.85)	(0.71)
Y_1	-2.940	-2.216	-3.963	-0.930	1.886	-0.406	-0.811	-2.586
	(4.08)	(3.59)	(2.66)	(2.69)	(3.02)	(0.97)	(1.39)	(1.45)
Y_2	2.668	2.689	6.133	1.174	-1.360	0.386	0.324	2.762
	(4.273)	(3.80)	(3.29)	(2.92)	(3.15)	(1.11)	(1.42)	(1.82)
Y_3	0.510	0.027	-0.766	0.434	-0.097	(nil)	0.180	0.135
	(0.80)	(0.80)	(0.72)	(0.59)	(0.21)		(0.17)	(0.13)
$w_1.w_2$	-0.063	-0.094	-0.403	0.353	0.166	-0.031	0.0368	0.102
	(0.48)	(0.458)	(0.28)	(0.34)	(0.40)	(0.24)	(0.33)	(0.27)
$Y_1.Y_2$	0.39595	0.627	0.402	0.764	-0.349	0.064	-0.494	-0.635
	(1.83)	(1.70)	(1.12)	(0.67)	(0.33)	(0.03)	(0.48)	(0.32)
$Y_1.Y_3$	0.30079	-0.144	0.028	0.504	0.044	0.0413	-0.005	-0.069
	(0.41)	(0.34)	(0.31)	(0.38)	(0.10)	(0.03)	(0.06)	(0.06)
$Y_2.Y_3$	-0.305	0.163	-0.015	-0.599	-0.028	-0.032	-0.8E-02	0.048
	(0.52)	(0.328)	(0.366)	(0.35)	(0.10)	(0.03)	(0.05)	(0.06)
$w_1.w_1$	-0.388	0.072	0.249	-0.329	0.845	0.479	0.248	-0.400
	(0.67)	(0.79)	(0.35)	(0.43)	(0.65)	(0.31)	(0.57)	(0.36)
$w_2.w_2$	-0.050	0.029	-0.179	2.2E-02	-0.237	(nil)	-0.076	-0.121
	(0.11)	(0.16)	(0.06)	(6 E-02)	(0.15)		(0.14)	(0.12)
$Y_1.Y_1$	-0.459	-0.532	-0.056	-0.432	0.157	(nil)	0.378	0.535
	(1.16)	(1.11)	(0.66)	(0.46)	(0.20)		(0.32)	(0.26)
$Y_2.Y_2$	0.106	-0.095	-0.353	-0.235	0.212	(nil)	0.240	0.158
	(0.79)	(0.63)	(0.54)	(0.31)	(0.21)		(0.22)	(0.13)
$Y_3.Y_3$	(nil)	0.7E-02	-0.001	-2.7E-05	-0.4E-03	(nil)	-0.060	-0.5E-03
		(0.02)	(0.01)	(0.01)	(1E-02)		(0.06)	(6E-03)
$w_1.Y_1$	2.59	3.499	1.321	1.2E-02	-0.731	0.217	0.286	0.565
	(1.77)	(1.97)	(1.34)	(0.89)	(1.18)	(0.53)	(0.80)	(0.57)
$w_1.Y_2$	-2.261	-3.286	-1.94	-5.2E-02	0.770	-0.242	-0.210	-0.483
	(1.63)	(1.88)	(1.37)	(0.93)	(1.18)	(0.54)	(0.70)	(0.57)
$w_1.Y_3$	-0.268	-0.165	0.387	1.7E-02	-0.4E-02	-0.030	-0.035	-0.008
	(0.28)	(0.43)	(0.19)	(0.23)	(0.08)	(0.04)	(0.06)	(0.04)
$w_2.Y_1$	-0.078	0.044	0.368	0.558	0.147	-0.213	0.108	0.266
	(0.81)	(0.50)	(0.34)	(0.54)	(0.314)	(0.34)	(0.43)	(0.25)
$w_2.Y_2$	0.101	-0.019	-0.231	-0.459	-0.098	0.209	-0.115	-0.126
	(0.815)	(0.46)	(0.32)	(0.50)	(0.34)	(0.34)	(0.43)	(0.22)
$w_2.Y_3$	-0.019	-0.077	-0.128	-0.128	0.3E-02	0.024	0.011	-0.032
	(0.16)	(0.20)	(0.14)	(0.19)	(0.03)	(0.02)	(0.03)	(0.03)

σ_u / σ_v	2.997 (2.28)	0.421 (41.88)	1.541 (2.27)	2.316 (2.21)	2.448 (1.65)	0.845 (1.36)	0.714 (1.77)	0.532 (2.08)
$\sqrt{\sigma_u^2 + \sigma_v^2}$	0.754 (0.48)	0.392 (7.82)	0.417 (0.70)	0.640 (0.61)	0.632 (0.12)	0.442 (0.28)	0.374 (0.26)	0.314 0.21

Table 4-6 (continued):

Variable	1987	1986	1985	1984	1983	1982	1981
Constant	16.976 (8.35)	10.734 (22.84)	14.907 (3.54)	3.659 (6.53)	5.436 (3.53)	1.120 (4.43)	5.162 (5.40)
w_1	-3.123 (2.01)	-2.296 (6.57)	-2.448 (1.48)	-0.192 (2.19)	-1.815 (1.59)	0.375 (1.20)	-0.870 (1.51)
w_2	1.063 (1.17)	0.176 (2.45)	-0.237 (0.574)	0.129 (0.90)	0.718 (0.83)	0.222 (0.83)	-0.179 (0.98)
Y_1	-0.398 (0.73)	0.255 (4.32)	0.642 (0.68	-0.549 (1.36)	1.410 (1.73)	0.313 (1.48)	-0.503 (1.33)
Y_2	-0.563 (1.13)	-0.491 (5.46)	-1.75 (0.67)	0.862 (1.57)	-1.027 (1.42)	0.296 (1.51)	0.670 (1.57)
Y_3	0.095 (0.19)	-0.082 (0.68)	(nil)	-0.133 (0.183)	-0.095 (0.13)	-0.207 (0.17)	-0.072 (0.19)
$w_1.w_2$	0.249 (0.37)	0.241 (0.83)	0.284 (0.29)	-0.015 (0.33)	-0.087 (0.31)	0.013 (0.27)	0.335 (0.34)
$Y_1.Y_2$	0.313 (0.39)	0.296 (0.80)	0.126 (0.04)	-0.126 (0.35)	-0.214 (0.46)	-0.258 (0.45)	-0.312 (0.21)
$Y_1.Y_3$	0.019 (0.03)	0.022 (0.21)	0.0936 (0.07)	-0.3E-03 (0.06)	0.063 (0.08)	0.044 (0.10)	-0.018 (0.07)
$Y_2.Y_3$	-0.024 (0.04)	-0.032 (0.23)	-0.108 (0.05)	0.010 (0.07)	-0.048 (0.08)	-0.020 (0.103)	0.024 (0.07)
$w_1.w_1$	0.388 (0.41)	0.477 (1.29)	0.437 (0.31)	-0.149 (0.41)	0.331 (0.44)	-0.120 (0.25)	0.028 (0.30)
$w_2.w_2$	0.139 (0.11)	0.008 (0.32)	-0.163 (0.12)	-0.077 (0.10)	0.104 (0.09)	-0.037 (0.12)	0.071 (0.13)
$Y_1.Y_1$	-0.137 (0.28)	-0.147 (0.67)	(nil)	0.165 (0.23)	0.066 (0.25)	0.168 (0.28)	0.202 (0.18)
$Y_2.Y_2$	-0.084 (0.13)	-0.082 (0.39)	(nil)	-0.082 (0.17)	0.174 (0.25)	0.110 (0.20)	0.155 (0.13)
$Y_3.Y_3$	0.002 (0.01)	0.018 (0.03)	(nil)	0.006 (0.01)	-0.9E-04 (0.01)	-0.003 (0.01)	-0.2E-02 (0.9E-02)
$w_1.Y_1$	0.477 (0.48)	0.235 (1.40)	-0.655 (0.40)	-0.026 (0.50)	-0.306 (0.72)	-0.252 (0.36)	0.221 (0.36)
$w_1.Y_2$	-0.176 (0.41)	-0.144 (1.30)	0.858 (0.37)	0.094 (0.53)	0.412 (0.76)	0.229 (0.41)	-0.104 (0.41)
$w_1.Y_3$	-0.024 (0.08)	0.027 (0.23)	0.017 (0.12)	0.026 (0.07)	0.007 (0.06)	0.038 (0.06)	0.044 (0.08)
$w_2.Y_1$	-0.326 (0.16)	-0.182 (0.59)	-0.123 (0.33)	-0.120 (0.21)	-0.051 (0.18)	-0.096 (0.23)	-0.197 (0.18)
$w_2.Y_2$	0.142 (0.08)	0.148 (0.45)	0.051 (0.35)	0.095 (0.21)	-0.005 (0.21)	0.06 (0.20)	0.193 (0.14)

$w_2.Y_3$	0.284 (0.03)	-0.048 (0.13)	-0.054 (0.05)	-0.017 (0.04)	0.037 (0.04)	-0.017 (0.04	-0.018 (0.05)
σ_u / σ_v	82.608 (2155)	126.29 (0.1E+05)	295.48 (0.3 E+05)	2.392 (1.23)	85.947 (2617)	1.155 (0.85)	1.239 (0.72)
$\sqrt{\sigma_u^2 + \sigma_v^2}$	0.656 (0.9E-01)	0.892 (0.12)	0.723 (0.6E-01)	0.369 (0.7E-01)	0.453 (0.5E-01)	0.332 (0.9E-01)	0.351 (0.7E-01)

Notes: (1) Standard errors of estimated parameters are shown in parentheses/ * significant with 95% confidence interval.
(2) Certain variables are excluded, due to high multicollinearity.

Table 4-7: Parameter estimates from *Model II* – Truncated distribution.

Variable	1995	1994	1993	1992	1991	1990	1989	1988
Constant	4.289	3.891	-7.08	-2.389	5.25	8.715	10.453	5.751
	(9.86)	(24.75)	(5.43)	(5.29)	(9.35)	(10.76)	(10.72)	(5.61)
w_1	0.256	-0.920	0.844	1.541	-2.747	-1.578	-1.416	-0.958
	(3.0)	(3.39)	(1.36)	(1.42)	(2.56)	(2.42)	(3.01)	(1.30)
w_2	-0.2E-02	0.453	0.203	-0.535	-1.156	0.063	-0.239	-0.864
	(1.2)	(1.71)	(1.20)	(1.03)	(1.32)	(0.98)	(0.94)	(0.72)
Y_1	-2.940	-2.216	-3.963	-0.864	1.886	-0.406	-0.810	-1.951
	(4.08)	(3.59)	(2.66)	(3.32)	(3.30)	(1.22)	(1.41)	(1.33)
Y_2	2.668	2.689	6.133	1.170	-1.3599	0.38580	0.32457	1.92846
	(4.27)	(3.80)	(3.26)	(3.32)	(3.450)	(1.409)	(1.429)	(1.4779)
Y_3	0.510	0.027	-0.766	0.419	-0.097	(nil)	0.180	0.166
	(0.80)	(0.80)	(0.72)	(0.60)	(0.22)		(0.17)	(0.11)
$w_1.w_2$	-0.063	-0.094	-0.402	0.364	0.166	-0.031	0.037	0.022
	(0.48)	(0.458)	(0.28)	(0.35)	(0.48)	(0.24)	(0.34)	(0.25)
$Y_1.Y_2$	0.395	0.627	0.401	0.796	-0.349	0.064	-0.444	-0.572
	(1.83)	(1.70)	(1.12)	(0.93)	(0.33)	(0.04)	(0.48)	(0.31)
$Y_1.Y_3$	0.301	-0.144	0.218	0.518	0.044	0.0413	-0.5E-02	-0.044
	(0.41)	(0.34)	(0.30)	(0.39)	(0.11)	(0.03)	(0.07)	(0.06)
$Y_2.Y_3$	-0.305	0.163	-0.148	-0.599	-0.028	-0.031	-0.8E-02	0.024
	(0.52)	(0.32)	(0.37)	(0.37)	(0.10)	(0.03)	(0.05)	(0.05)
$w_1.w_1$	-0.388	0.007	0.249	-0.344	0.845	0.478	0.246	(nil)
	(0.67)	(0.79)	(0.35)	(0.44)	(0.72)	(0.50)	(0.57)	
$w_2.w_2$	-0.050	0.3 E-03	-0.179	0.17	-0.237	(nil)	-0.075	-0.104
	(0.10)	(0.16)	(0.05)	(0.06)	(0.15)		(0.14)	(0.13)
$Y_1.Y_1$	-0.458	-0.532	-0.556	-0.458	0.157	(nil)	0.301	0.461
	(1.16)	(1.11)	(0.66)	(0.52)	(0.20)		(0.33)	(0.24)
$Y_2.Y_2$	0.106	-0.095	-0.353	-0.251	0.212	(nil)	0.240	0.175
	(0.80)	(0.63)	(0.54)	(0.52)	(0.21)		(0.22)	(0.11)
$Y_3.Y_3$	(nil)	0.7 E-02	-0.001	-3.6E-04	-0.5 E-03	(nil)	-0.6E-02	(nil)
		(0.01)	(0.01)	(0.01)	(0.01)		(0.01)	
$w_1.Y_1$	2.589	3.499	1.321	2.2E-02	-0.733	0.217	0.286	0.433
	(1.77)	(1.97)	(1.38)	(1.11)	(1.31)	(0.69)	(0.80)	(0.57)
$w_1.Y_2$	-2.261	-3.286	-1.940	-4.1E-02	0.769	-0.242	-0.210	-0.305
	(1.63)	(1.88)	(1.36)	(1.15)	(1.30)	(0.69)	(0.70)	(0.53)
$w_1.Y_3$	-0.268	-0.165	0.387	6.1E-03	-0.4E-02	-0.029	-0.036	-0.02
	(0.28)	(0.43)	(0.19)	(0.23)	(0.08)	(0.05)	(0.07)	(0.04)
$w_2.Y_1$	-0.07	0.044	0.368	0.605	0.147	-0.213	0.107	0.231
	(0.81)	(0.50)	(0.34)	(0.56)	(0.31)	(0.406)	(0.46)	(0.24)
$w_2.Y_2$	0.101	-0.019	-0.231	-0.506	-0.097	0.209	-0.114	-0.13
	(0.81)	(0.46)	(0.31)	(0.52)	(0.34)	(0.39)	(0.45)	(0.22)
$w_2.Y_3$	-0.019	-0.077	-0.128	-0.135	0.4 E-02	0.025	0.011	-0.023
	(0.16)	(0.20)	(0.11)	(0.19)	(0.03)	(0.03)	(0.03)	(0.03)
μ / σ_u	0.1E-05	-5E-09	-0.1E-04	0.7E-04	-0.2E-06	0.2E-07	-0.3E-07	-0.371
	(2.57)	(608.4)	(9.27)	(5.13)	(3.50)	(84.97)	(86.09)	(318.7)
σ_u / σ_v	2.997	0.421	1.541	1.871	2.448	0.846	0.714	0.478

	(2.28)	(41.88)	(2.27)	(2.16)	(2.26)	(12.97)	(8.48)	(25.27)
$\sqrt{\sigma_u^2 + \sigma_v^2}$	0.754 (0.48)	0.392 (7.82)	0.417 (0.70)	0.522 (0.510)	0.632 (0.49)	0.442 (3.84)	0.374 (2.30)	0.311 (3.85)

Table 4-7 (continued):

Variable	**1987**	**1986**	**1985**	**1984**	**1983**	**1982**	**1981**
Constant	11.614 (6.39)	5.522 (9.42)	7.180 (6.73)	3.121 (5.81)	5.925 (4.54)	1.41 (4.44)	4.736 (5.13)
w_1	-2.193 (1.76)	-1.608 (2.66)	-1.584 (1.82)	0.199 (1.93)	-1.908 (1.77)	0.539 (1.28)	-0.743 (1.4)
w_2	0.341 (1.02)	-0.198 (1.62)	0.254 (1.36)	0.099 (0.79)	0.854 (1.26)	0.240 (0.85)	-0.148 (1.02)
Y_1	-0.671 (1.06)	-0.321 (2.31)	1.462 (1.80)	-0.551 (1.05)	1.471 (1.43)	0.204 (1.55)	-0.517 (1.33)
Y_2	0.231 (1.29)	0.716 (2.54)	-1.334 (1.61)	0.911 (1.19)	-1.149 (1.15)	0.463 (1.62)	0.735 (1.54)
Y_3	0.084 (0.16	-0.175 (0.25)	-0.193 (0.28)	-0.143 (0.17)	-0.072 (0.19)	-0.214 (0.17)	-0.081 (0.17)
$w_1.w_2$	0.155 (0.39)	0.019 (0.60)	-0.083 (0.41)	-0.009 (0.27)	-0.104 (0.39)	0.001 (0.20)	0.322 (0.33)
$Y_1.Y_2$	0.055 (0.32)	0.012 (0.45)	0.191 (0.22)	-0.132 (0.21)	-0.255 (0.54)	-0.276 (0.46)	-0.313 (0.22)
$Y_1.Y_3$	-0.012 (0.04)	0.006 (0.09)	0.133 (0.11)	0.9E-03 (0.05)	0.061 (0.07)	0.036 (0.10)	-0.020 (0.06)
$Y_2.Y_3$	0.005 (0.04)	0.006 (0.101)	-0.113 (0.08)	0.010 (0.06)	-0.050 (0.06)	-0.011 (0.11)	0.026 (0.07)
$w_1.w_1$	0.249 (0.44)	0.162 (0.54)	0.351 (0.46)	-0.191 (0.35)	0.354 (0.48)	-0.161 (0.27)	0.004 (0.29)
$w_2.w_2$	0.016 (0.11)	-0.051 (0.19)	-0.171 (0.14)	-0.079 (0.08)	0.107 (0.09)	-0.051 (0.13)	0.068 (0.13)
$Y_1.Y_1$	0.004 (0.22)	0.045 (0.35)	(nil)	0.168 (0.14)	0.089 (0.26)	0.184 (0.30)	0.203 (0.18)
$Y_2.Y_2$	0.010 (0.12)	-0.036 (0.19)	-0.149 (0.19)	-0.007 (0.13)	0.197 (0.29)	0.107 (0.19)	0.152 (0.13)
$Y_3.Y_3$	0.004 (0.01)	0.007 (0.01)	(nil)	0.007 (0.01)	-0.5E-03 (0.5E-02)	-0.002 (0.01)	-0.003 (0.01)
$w_1.Y_1$	0.512 (0.58)	0.263 (0.70)	-1.27 (0.78)	-0.017 (0.34)	-0.365 (0.74)	-0.232 (0.36)	0.229 (0.35)
$w_1.Y_2$	-0.293 (0.54)	-0.155 (0.68)	1.351 (0.80)	0.075 (0.35)	0.472 (0.73)	0.120 (0.42)	-0.124 0.50
$w_1.Y_3$	-0.027 (0.06)	0.029 (0.13)	-0.002 (0.13)	0.050 (0.05)	0.011 (0.07)	0.040 (0.06)	-0.124 (0.40)
$w_2.Y_1$	-0.257 (0.24)	-0.067 (0.46)	-0.043 (0.24)	-0.103 (0.13)	-0.054 (0.22)	-0.098 (0.22)	0.045 (0.07)
$w_2.Y_2$	0.182 (0.22)	0.106 (0.42)	-0.007 (0.22)	0.082 (0.13)	-0.021 (0.24)	0.062 (0.19)	-0.195 (0.19)
$w_2.Y_3$	0.007	-0.044	-0.034	-0.020	0.008	-0.017	0.188

	(0.03)	(0.07)	(0.09)	(0.03)	(0.05)	(0.04)	(0.14)
μ / σ_u	-3.168 (16.16)	-3.357 (20.40)	-0.754 (13.03)	-1.094 (4.71)	0.6E-01 (0.65)	-0.611 (18.97)	-0.3E-06 (8.35)
σ_u / σ_v	13.090 (24.56)	13.487 (26.50)	1.220 (3.18)	2.7668 (2.14)	104.30 (2600)	1.095 (2.00)	1.226 (1.64)
$\sqrt{\sigma_u^2 + \sigma_v^2}$	1.544 (3.20)	2.226 (5.36)	0.506 (0.83)	0.462 (0.39)	0.458 (0.12)	0.342 (0.67)	0.348 (0.37)

Notes: (1) Standard errors of estimated parameters are shown in parentheses/ * significant with 95% confidence interval.
(2) Certain variables are excluded, due to high multicollinearity.

<u>Table 4-8:</u> Parameter estimates from *Model III* – Exponential distribution.

Variable	**1995**	**1994**	**1993**	**1992**	**1991**	**1990**	**1989**	**1988**
Constant	4.518	5.064	-6.969	-2.24	6.082	9.96	10.51	5.750
	(8.62)	(5.48)	(4.73)	(4.33)	(9.56)	(7.85)	(9.78)	(5.72)
w_1	0.256	-1.064	0.844	1.541	-2.596	-1.724	-1.418	-0.958
	(2.63)	(1.37)	(1.26)	(1.20)	(2.45)	(2.32)	(3.00)	(1.31)
w_2	-0.2E-02	0.304	0.203	-0.535	-0.535	-0.025	-0.239	-0.863
	(1.21)	(1.01)	(1.09)	(0.94)	(1.19)	(0.94)	(0.81)	(0.73)
Y_1	-2.940	-2.511	-3.963	-0.863	1.185	-0.180	-0.811	-1.951
	(2.63)	(2.15)	(2.26)	(2.48)	(2.64)	(1.25)	(1.40)	(1.35)
Y_2	2.667	2.677	6.132	1.170	-0.682	-0.012	0.324	1.928
	(2.88)	(2.28)	(2.56)	(2.65)	(2.59)	(1.40)	(1.41)	(1.49)
Y_3	0.510	0.246	-0.766	0.418	-0.064	(nil)	0.180	0.166
	(0.62)	(0.49)	(0.72)	(0.52)	(0.22)		(0.16)	(0.11)
$w_1.w_2$	-0.063	-0.078	-0.402	0.364	0.017	0.032	0.037	0.022
	(0.48)	(0.34)	(0.25)	(0.30)	(0.403)	(0.22)	(0.325)	(0.25)
$Y_1.Y_2$	0.395	0.879	0.401	0.796	-0.354	0.077	-0.443	-0.572
	(1.02)	(0.95)	(1.08)	(0.72)	(0.23)	(0.03)	(0.48)	(0.32)
$Y_1.Y_3$	0.300	-0.216	0.022	0.517	0.024	0.033	-0.5E-02	-0.043
	(0.28)	(0.21)	(0.27)	(0.36)	(0.08)	(0.03)	(0.06)	(0.05)
$Y_2.Y_3$	-0.305	0.212	-0.015	-0.599	-0.012	-0.029	-0.7E-02	0.024
	(0.34)	(0.19)	(0.33)	(0.34)	(0.08)	(0.02)	(0.05)	(0.05)
$w_1.w_1$	-0.388	0.119	0.249	-0.344	0.791	0.486	0.248	(nil)
	(0.61)	(0.579)	(0.34)	(0.40)	(0.53)	(0.48)	(0.56)	
$w_2.w_2$	-0.050	0.091	-0.179	1.7E-02	-0.163	(nil)	-0.076	-0.104
	(0.09)	(0.11)	(0.053)	(0.05)	(0.13)		(0.14)	(0.13)
$Y_1.Y_1$	-0.458	-0.675	-0.055	-0.458	0.163	(nil)	0.379	0.461
	(0.65)	(0.61)	(0.62)	(0.44)	(0.164)		(0.323)	(0.24)
$Y_2.Y_2$	0.106	-0.168	-0.353	-0.251	0.219	(nil)	0.240	0.175
	(0.52)	(0.35)	(0.51)	(0.37)	(0.1)		(0.22)	(0.11)
$Y_3.Y_3$	2.590	0.009	-0.1E-02	-6.6E-04	-0.3E-03	(nil)	-0.6E-02	(nil)
	(1.57)	(0.01)	(0.01)	(0.01)	(0.8E-02)		(0.6E-02)	
$w_1.Y_1$	-2.261	3.959	1.321	2.3E-02	-0.301	0.127	0.286	-0.305
	(1.45)	(1.05)	(1.19)	(0.88)	(1.11)	(0.71)	(0.81)	(0.57)
$w_1.Y_2$	-0.268	-3.770	-1.94	-4.1E-02	0.292	-0.132	-0.210	-0.027
	(0.24)	(1.03)	(1.15)	(0.89)	(1.09)	(0.70)	(0.71)	(0.03)
$w_1.Y_3$	-0.078	-0.191	0.387	6.3E-03	-0.013	-0.021	-0.036	-0.131
	(0.78)	(0.19)	(0.19)	(0.21)	(0.07)	(0.05)	(0.06)	(0.22)
$w_2.Y_1$	0.102	-0.148	0.368	0.605	0.147	-0.042	0.108	0.231
	(0.77)	(0.37)	(0.33)	(0.47)	(0.26)	(0.39)	(0.44)	(0.24)
$w_2.Y_2$	-0.019	0.159	-0.231	-0.507	-0.129	0.440	-0.115	-0.132
	(0.12)	(0.34)	(0.30)	(0.45)	(0.31)	(0.39)	(0.44)	(0.22)
$w_2.Y_3$	(nil)	0.2E-03	-0.128	-0.135	-0.4E-03	(nil)	0.011	-0.023
		(0.13)	(0.13)	(0.16)	(0.04)		(0.03)	(0.03)
θ	2.926	2.039	5.986	4.547	2.952	5.853	9.638	4.987
	(1.21)	(0.32)	(5.17)	(2.90)	(0.77)	(10.77)	(30.02)	(5.68)
σ_v	0.355	2.7E-06	0.261	0.299	0.25966	0.344	0.315	0.232
	(0.11)	(7E-02)	(0.8E-01)	(0.9E-01)	(0.9E-01)	(0.12)	(0.11)	(0.13)

Table 4-8 (continued):

Variable	1987	1986	1985	1984	1983	1982	1981
Constant	10.984 (4.19)	8.506 (8.70)	9.372 (4.81)	3.967 (5.83)	7.469 (3.21)	1.646 (4.06)	5.376 (5.01)
W_1	-2.406 (1.17)	-1.831 (2.45)	-0.573 (1.18)	-0.219 (1.95)	-2.592 (0.96)	0.244 (1.06)	-0.908 (1.37)
W_2	0.2E-03 (0.60)	0.180 (1.35)	-1.294 (0.97)	0.137 (0.79)	0.920 (0.64)	0.213 (0.75)	-0.194 (0.92)
Y_1	-0.094 (0.91)	0.230 (1.36)	0.840 (0.56)	-0.425 (1.26)	1.527 (1.16)	0.409 (1.35)	-0.521 (1.22)
Y_2	-0.283 (1.14)	-0.112 (1.81)	-1.52 (0.56)	0.705 (1.42)	-1.311 (1.14)	0.148 (1.38)	0.673 (1.43)
Y_3	0.014 (0.09)	-0.2140 (0.19)	-0.222 (0.13)	-0.124 (0.16)	-0.020 (0.11)	-0.200 (0.16)	-0.072 (0.18)
$w_1.w_2$	0.143 (0.22)	0.033 (0.42)	0.646 (0.31)	-0.016 (0.28)	-0.202 (0.19)	0.028 (0.23)	0.344 (0.32)
$Y_1.Y_2$	-0.055 (0.19)	-0.048 (0.36)	0.120 (0.04)	-0.111 (0.29)	-0.361 (0.30)	-0.278 (0.41)	-0.314 0.18
$Y_1.Y_3$	0.831 (0.04)	0.047 (0.06)	0.075 (0.04)	0.005 (0.06)	0.055 (0.05)	0.044 (0.09)	-0.315 (0.18)
$Y_2.Y_3$	-0.009 (0.04)	-0.018 (0.06)	-0.081 (0.03)	0.004 (0.06)	-0.049 (0.05)	-0.021 (0.10)	-0.019 (0.06)
$w_1.w_1$	0.395 (0.20)	0.087 (0.51)	0.154 (0.30)	-0.169 (0.34)	0.483 (0.24)	-0.084 (0.21)	0.025 (0.07)
$w_2.w_2$	-0.020 (0.07)	0.196 (0.16)	-0.118 (0.08)	-0.080 (0.08)	0.071 (0.10)	-0.026 (0.10)	0.042 (0.27)
$Y_1.Y_1$	0.049 (0.16)	0.021 (0.28)	(nil)	0.140 (0.21)	0.135 (0.17)	0.172 (0.26)	0.075 (0.12)
$Y_2.Y_2$	0.0795 (0.06)	0.056 (0.10)	(nil)	0.004 (0.15)	0.261 (0.17)	0.129 (0.18)	0.203 (0.16)
$Y_3.Y_3$	-0.2E-02 (0.5E-02)	0.176 (0.54)	(nil)	0.006 (0.01)	-0.3E-03 (0.7E-02)	-0.003 (0.01)	0.158 (0.11)
$w_1.Y_1$	0.241 (0.45)	0.029 (0.54)	-0.656 (0.35)	0.015 (0.40)	-0.374 (0.34)	-0.280 (0.32)	-0.003 (0.01)
$w_1.Y_2$	-0.047 (0.45)	-0.014 (0.06)	0.706 (0.34)	0.064 (0.43)	0.520 (0.33)	0.264 (0.37)	0.235 (0.33)
$w_1.Y_3$	-0.018 (0.04)	-0.067 (0.19)	0.097 (0.05)	0.027 (0.06)	-0.003 (0.04)	0.037 (0.06)	-0.119 (0.37)
$w_2.Y_1$	-0.166 (0.11)	0.156 (0.16)	-0.074 (0.13)	-0.123 (0.20)	-0.079 (0.12)	-0.093 (0.21)	0.045 (0.07)
$w_2.Y_2$	0.132 (0.11)	0.158 (0.17)	0.125 (0.12)	0.096 (0.19)	-0.002 (0.12)	0.062 (0.19)	-0.194 (0.15)
$w_2.Y_3$	-0.013 (0.02)	-0.048 (0.04)	-0.102 (0.03)	-0.017 (0.03)	0.015 (0.027)	-0.018 (0.04)	0.192 (0.12)
θ	2.257	1.971	2.197	5.805	2.841	6.825	7.013

	(0.34)	(0.32)	(0.34)	(1.96)	(0.51)	(3.16)	(3.31)
σ_v	0.1E-04 (0.4E-01)	0.2E-05 (2.85)	0.6E-06 (0.6E-01)	0.169 (0.4E-01)	0.2E-04 (0.5E-01)	0.217 (0.5E-01)	0.232 (0.6E-01)

Notes: (1) Standard errors of estimated parameters are shown in parentheses/ * significant with 95% confidence interval.

(2) Certain variables are excluded, due to high multicollinearity.

CHAPTER 5: Market Structure and Profit Relationship in Banking

5.1 Introduction

The primary objective of the current chapter is to examine the relationship between market structure and profit the UK banking industry. Many studies in the banking literature (especially in the US) examine the connection between profitability and measures of market structure (concentration or market share) and find a positive statistical relationship. There are two different interpretations of this phenomenon: the first is *the market power hypothesis*, and the second is *the efficient structure hypothesis*. The *market-power* explanation suggests that firms in concentrated markets exercise market power in pricing and earn supernormal profits [Shepherd (1982)]. The *efficient structure* paradigm links concentration to high profitability through efficiency [Demsetz (1973)]. The main problem of the literature has been the interpretation of the positive relationship between profitability and market structure, and whether this supports the market power or the efficient structure hypothesis due to the exclusion of a direct efficiency measure [Berger (1995a, 1997), Berger and Hannan (1998)]. To overcome the problem of most of the previous literature, direct measures of both market structure and efficiency are included.

Following deregulation, banking industry is consolidating rapidly. Hence, it is important to understand which market structure best serves the deposit and credit needs of an economy. The different explanations of differences in profitability across banks have direct opposing implications for antitrust policy. If high profits are created by market power, then antitrust actions are likely to be socially beneficial by moving prices towards competitive levels and allocating resources more effectively. On the other hand, if high efficiency is the reason for high profits, then antitrust actions such as breaking-up efficient firms that have gained large market share, or disallow efficient firms to acquire other firms, will probably raise costs and lead to prices less favourable to the consumers [Berger (1995a, 1997), Berger and Humphrey (1997)].

The structure of this chapter is the following: section 2 distinguishes between the four hypotheses, namely the two market-power hypotheses, and the two efficient-structure hypotheses, and gives an overview of the different performance and market structure variables utilised in the literature. Section 3 reviews the literature. Section 4 describes the methodology, by giving the structural models underlying the various hypotheses, outlines the equations to be estimated in hypothesis tests, and provides a brief description of the efficiency measures. Section 5 describes the data set employed. Section 6 summarises the empirical results, and section 7 concludes.

5.2 The theory behind the relevant hypotheses

There are two different interpretations for the positive statistical relationship between profitability and market structure found in the literature: the *market power* hypothesis [Shepherd (1982), Clark (1986), Berger and Hannan (1989), Golberg and Rai (1995)], and the *efficient structure* hypothesis [Smirlock (1985), Evanoff and Fortier (1988), Jackson (1992)].

The first one, the *market power* hypothesis (MP), states that banks in concentrated markets are able to extract monopolistic profits by their ability to offer lower deposit rates and charge higher interest on loans [Berger and Hannan (1993), Berger (1995a, 1997), Goldberg and Rai (1995)]. The theory is divided into two hypotheses. The *structure conduct performance* (SCP) hypothesis states that the positive relationship between profitability and market structure reflect the setting of prices that are less favourable to consumers, i.e. lower deposit rates higher loan rates, in more concentrated markets due to competitive imperfections. This hypothesis is based on the model of oligopolistic behaviour of firms, where collusive arrangements are less costly to maintain in a concentrated market.

The *relative market hypothesis* (RMP), which is related to SCP, states that only firms with large market shares and well differentiated products are able to exercise market power in pricing of these products and earn supernormal profits. The difference between the SCP and RMP is that the latter does not assume concentrated markets, and moreover, anti-competitive advantages due to size can exist non-concentrated.

On the other hand, according to the *efficient structure* hypotheses (ES), efficient firms increase in size and market share because of their ability to generate higher profits that, in turn, usually lead to higher market concentration [see Demsetz (1973)]. This hypothesis suggests that both profitability and market structure reflect important efficiency differences across firms. ES is divided into two sub-hypotheses. The first one is the *X-efficiency* (ESX) hypothesis, according to which firms earn abnormal profits steaming from lower costs due to superior management or production technologies. The second is the *scale efficiency* (ESS), which states that firms have essentially equally good management and technology, but some firms produce at a more efficient scale than others do. Consequently, they obtain lower unit cost and higher market profits [Smirlock (1985), Berger and Hannan (1993), Berger (1995a, 1997), Goldberg and Rai (1995)].

There is an important public policy question on which of these two hypotheses represents the market. If the first hypothesis is valid, then motivation for mergers may arise from the desire to set unfavourable price for consumers, which will have as a result the decrease in the consumer and the producer surplus. On the other hand, if the ES hypothesis is valid, then mergers may be motivated by efficiency considerations that would increase total surplus. The first category suggests that antitrust or regulatory action may be productive, whereas the second category suggests that such actions are likely to be counterproductive and socially costly [Berger and Hannan (1993, 1998)].

5.2.1 Defining Bank's Performance

Previous studies have mainly used two types of measures for the evaluation of banks performance. The first is related to the price of a particular product or service, while the other one uses profitability measures. A survey of 73 US studies by Molyneux *et al.* (1996) indicates that the most commonly used price measures are the average interest paid on loans, the average interest paid on deposits, and the average interest charge on demand deposits [e.g Berger and Hannan (1989, 1992), Calem and Carlino (1991), Jackson (1992), Beger and Hannan (1993, 1998)]. However, the use of price measures has been criticised on the grounds that price measures combine flow variables with stock variables.[59] Furthermore, studies that use price as an indicator of bank's performance have a very small adjusted R^2 (always below 0.50). As Gilbert (1984) points out '*the estimated effects of market concentration on the average interest rates on loans may be biased if the measure of market concentration is correlated with the omitted variables*'[60].

The other category is the profitability measures. There are two major advantages for the use of profitability measures. The first one is that by use of a single number one can consolidate information about a multiproduct firm; by definition, banks are multiproduct entities. And second, they are very simple to calculate. Nevertheless, they face the same drawback as the first category, as they combine flow variables with stock variables. The most commonly used profitability measures are return on assets (ROA) and return of capital employed (ROE) [e.g Smirlock (1985), Evanoff and Fortier (1988), Clark (1986), Thorton and Molyneux (1992), Berger (1995a,1997)].

5.2.2 Concentration Measures

To account for market structure, most US studies have used different concentration measures. The most commonly used are the concentration ratio and the Herfindahl index. The *concentration ratio* is the proportion of output produced by the top *n* firms in the industry [e.g Rhodes (1980,1981), Rhodes and Savage (1981), Marlow (1982), Clark (1986), Jakson (1992), Thorton and Molyneux (1992), Goldberg and Rai (1995), Berger (1997)]. The major criticism of the above measure is that it measures only one point on the concentration curve[61] (since it is derived directly from the concentration curve), and so the ranking of industries depends critically on the points chosen. Further, it does not take into account the dispersion of bank size in the market and, also, it does not reflect the number of competing companies.

Hannah and Kay (1976) have introduced a measure that takes into account all firms in the market [e.g Kwast and Rose (1984), Hannan (1984), Goldberg and Rai (1995),

[59] For example, they use interest on loans over one period, which is a flow variable, with loans due at the end of the year, which a stock variable.
[60] Gilbert (1984) p.p. 631.
[61] The concentration curve is constructed by plotting the cumulative shares of a market output to the cumulative number of firms from the larger to the smallest.

Berger (1995a, 1997)]. This measure is an asymmetric strictly concave function and its simplest form is given by:

$$R = \sum_i S_i^{\alpha} \tag{5.1}$$

where S_i^{α} is the market share of firm i, and α is the elasticity parameter, i.e. the value which determines the weights given to large firms relative to small ones. For example, as α tends to zero, the index is simply the numbers of the firms; on the other hand, as α becomes larger, the weight given to small firms becomes negligible. The *Herfindahl index*, has a value of α =2 and R is given by:

$$R_H = \sum_i S_i^2 \tag{5.2}$$

This measured is considered superior compared to the first one, because it overcomes its problems, as reflects the number and dispersion of firms in the market [Molyneux *et al.* (1996)]. Nonetheless, empirical studies find that these two measures are so highly correlated and that the choice of measure is not of critical importance for hypothesis testing [see Rose and Fraser (1976), and Heggestad (1979)].

5.2.3 Market Share

Most studies use a weight to measure the market share of each bank, which is usually the proportion of each individual bank's deposit in the market relative to the total market deposit [e.g Smirlock (1985), Berger and Hannan (1989,1992,1993,1998), Jakson (1992), Berger (1995a, 1997), Goldberg and Rai (1995)]

5.2.4 Control Variables

The major control variables that have been utilised from previous studies are the following:

1. The *size variable* is the most common variable used to account for cost d+ifferences across banks and to control for the ability of larger banks to diversify, and in most cases it is the amount of total assets of each individual bank. Cost differences related to bank size will lead to a positive coefficient if there are

economies of scales; a negative coefficient implies that increased diversification leads to lower risk, and thus lower required returns. In some studies the log of the total assets is used to reduce the scale effect. [e.g Clark (1986), Goldberg and Rai (1995)]

2. The *wage rates*[62] is used to capture the effect of wages and, in turn, of operating cost, on profits. It is expected to be negative when the efficient structure hypothesis holds, because efficient banks are expected to operate at lower costs. The same argument can be applied if the market power hypothesis holds. [e.g Thorton and Molyneux (1992), Goldberg and Rai (1995)].
3. A variable used to account for different risk levels between banks is the *loan to asset ratio*. This is used as a proxy of portfolio risk based on the view that loans tend to be risky relative to other assets held in bank's portfolio. A lower ratio indicates a more risky position [e.g Smirlock (1985), Goldberg and Rai (1995)].
4. Most US studies include a variable for *entry barriers* due to the nature of the US banking industry usually measured by total market deposits. Entry conditions are included in the structure performance models in order to evaluate their impact on bank performance and to see their relationship with to concentration. The majority of studies that account for this variable assume that lower entry barriers (the ability to undertake branching) enter the performance equation only as a shift parameter. The literature suggests that higher entry barriers lead to higher profitability [e.g. Rhoades (1980,1981,1982a), Rhoades and Rutz (1982), Smirlock (1985), Berger and Hannan (1989)].

[62] Wage rate is usually proxy for the cost of labour, and is calculated as the ratio of total wages and salaries to the number of employees.

5.3 Literature Review

Four approaches have been considered in the literature to distinguish among the different hypotheses, and each has its benefits and limitations. The first one, regresses profits on the market structure variables, concentration and market share. Studies that utilised this method find a statistically significant positive relationship between profit and market share, whereas the coefficient of concentration is found to be insignificant [see Table 5-1, Gilbert (1984), and Berger and Hannan (1993)]. The limitation of this approach lies in the interpretation of the findings. Some argue that these findings constitute a support for the ES hypothesis based on the notion that both profits and market share are correlated with efficiency, which is excluded from the empirical specification [Smirlock, Gillian, and Marshall (1984,1986), Smirlock (1985)]. Others state that the findings support the MP hypothesis since firms with larger shares can exercise greater market power and earn higher profits [Shepherd (1986)].

The second approach tries to overcome this problem by including a direct measure of efficiency in the profitability equation. According to this approach, if efficiency is properly controlled for, then the market share coefficient should reflect solely market power effects. Some studies include a proxy of scale efficiency to reflect the scale version of the ES hypothesis [Allen and Hagin (1989)]. More recent studies have also included explicit measures of X-efficiency in the profitability equation to control for the ESX hypothesis [Timme and Yang (1991), Berger (1995a)].

The major drawback of the two approaches mentioned above is that the dependent variable, profitability, may include a considerable amount of noise. A portion of this noise may be generated by difficulties in using accounting data [Fisher and McGowen (1983)]. Moreover, profits may be affected by other factors such as loan loss provisions, which are unrelated to either structure and cost efficiency [Berger and Hannan (1993)].

The third approach emerges from the need to overcome the problem due to the use of profit as the dependent variable. Instead of profit data, some studies utilise more precise information on the prices of individual bank deposit and loan products [Berger and Hannan (1989), Hannan (1991), Jakson (1992), Berger and Hannan (1992)]. These variables are regressed on the market share variables to test the MP hypothesis. With lower deposit rates and/or higher loan rates, i.e. prices less favourable to consumers, the MP hypotheses is supported. The major advantage of this method is that exact prices (paid or received) are more accurate indicators of the market power than profits. On the other hand, it may be considered as problematic due to the exclusion of efficiency measures in the analysis.

The last approach, which has been applied only by a handful of new studies, distinguishes among all the four hypotheses, SCP, RMP, ESX, and ESS, by using direct measures for both market structure and efficiency. This approach is flexible in the sense that all four hypotheses are represented by different variables, so that any (or all of them) may be found to be consistent with the data employed [Berger and Hannan (1993,1998), Berger (1995a,1997), Goldberg and Rai (1996)].

These studies directly relate market structure to efficiency. Concentration and market share are regressed on the efficiency measures to test the efficient structure condition that efficiency creates greater concentration or market share [see Berger (1995a,1997)]. '*For the efficient structure hypothesis to determine the positive relationship between performance and structure spuriously, efficiency must be positively related to both performance and structure. Under the MP hypotheses, market structure is associated with market power, and firms may take some of the benefits of this power as a more relaxed environment in which there is less pressure to maximise efficiency*'.[63] As suggested by Hicks (1935), the reduction of competitive pressure in concentrated markets may result in higher cost per unit of output because of slack management. Under this 'quiet life' addition to the MP hypotheses, higher concentration and market share may be negatively related to efficiency.

[63] Berger (1997), p.p. 8.

5.3.1 US empirical evidence

Early literature concentrated on testing of *SCP hypothesis*, based on the notion that the positive statistical relationship between market concentration and measures of firm or industry profitability reflects the setting of prices; the latter are unfavourable to consumers in more concentrated markets as a result of collusion or other forms of non-competitive behaviour. Demzet (1973, 1974) and Peltzman (1977) question this assumption with the introduction of the efficient structure hypothesis. They stated that firms possessing a comparative advantage in production become larger and, hence, they gain a higher market share. This dispersion of efficiency within markets creates higher levels of concentration and results in a greater than average efficiency within these markets, which in turns yields a positive profit concentration relationship. Consequently, high concentration is due to the high market share of efficient firms. Thus market share proxies for relative efficiency and therefore will be positively correlated with profitability.

Gilbert (1984) provides a survey of forty-four empirical studies that examine the relationship of market structure with the performance of banking institutions. From these studies thirty-two report some evidence of significant association between market structure and measures of bank performance, with the direction of influence of market structure as indicated by the SCP hypothesis. In seven studies the coefficients of the measures of market structure are found to be insignificant. However, these studies face important limitations. First, average interest and service charge rates are poor measures of bank performance[64], whereas bank profits are more appropriate measure of the performance. Second, the effect of regulation on bank's performance is missing from most of these studies[65]. Since there is no theoretical foundation for the effects of regulation on the relationship tested, the results can not provide any

[64] (1) For a period of time due to the effect of Regulation Q, the deposit ceiling rates were below the market rate, so the average interest paid on time and saving deposits does not represents markets structure, but most likely a function of the maturity distribution of bank's deposits. (2) For all those measures the numerator is an annual income or expense flow, where the denominator is a balance sheet item recorded as of a point in time.

[65] Gilbert states that changes in the regulation do affect the relationship. He presents as an example the influence of entry regulation. If this is controlled by the regulators, the firms already in the market determine to what extent the pricing and the availability of services reflects monopolistic and competitive behaviour. On the other hand, in the case of unregulated environment, pricing of banking

information for the influence of changes in the regulatory system on structure performance relationships [Gilbert (1984)]. Third, there is great variation in the measure of market structure and bank performance. Studies that use the most appropriate measures for bank performance find significant influence of market structure on these performance measures. However, statistical results from most of the studies are very small. Finally, these studies do not include a market share variable. Hence, the effect of concentration reported in these studies is spurious and most probably due to the correlation between profitability and the omitted market share variable [Smirlock (1985)]. Table 5-1 provides a summary of the early studies together with their conclusions.

services is influenced by the threat of entry by new firms, irrespective of the existing structure of the industry.

Table 5-1: Review of the early literature on market structure and performance relationship[66]

Authors	Measure of Bank performance	R^2 or $\overline{R}^2$	Measure of Market Structure	Are the coefficients on the Market Structure Significant
Edwards (1964)	Interest rates on business loans	0.36-0.64	CR3	Yes for 1955 No for 1957
Fleching (1965)	Interest rates on business loans	0.16-0.48	CR3	No
Meyer (1967)	Interest rates on business loans	0.67-0.73	CR3	Yes for 1955 No for 1957
Philip (1967)	Interest rates on business loans	0.51-0.64	CR3	Yes
Bell and Murphy (1969)	Estimated service charge on demand deposits	0.22-029	CR3	Yes
Brucker (1970)	Elasticity of loans demand	0.57	CR3	Yes
Fraser and Rose (1971)	1. IL/TL 2. IT/TS 3. SC/DD 4. NI/C	0.41-0.54 0.03-0.14 0.21-.030 0.07-0.15	CR1	1. No for 1966 Yes for 1967 2. No 3. No 4. No
Vernor (1971)	NI/C	0.21	CR3	Yes
Jacobs (1971)	Interest rates on business loans	0.18-0.25	CR3	Yes
Ware (1972)	SC/DD NI/C IL/TL IT/TS	0.49-0.51 0.26-0.45 0.42-0.43 0.49-0.61	CR2	No No No No
Edwards (1973)	NI/C	0.05	CR3	No
Yates (1974)	IL/TL IT/TD NI/TA Portfolio selection	0.14-.035	H Changes in H Market share stability	Small effects of concentration, important effect for changes in concentration

[66] Sources: (1.) Gilbert, A. "Bank Market Structure and Competition", Journal of Money, Credit and Banking, Vol. 16, No. 4, November 1984, p.p. 617-645, (2.) Molyneux,P., Altunbas,Y and Gardener,E., "Efficiency in European Banking", USA: John Wiley & Sons, Inc., 1996, (3) Individual articles.

Fraser and Alvis (1975)	NI/TA NI/C IL/TL SC/DD IT/TS	—	Dummy variables for markets with relatively high CR1	No for all performance variables
Fraser and Rose (1976)	IL/TL IT/TS SC/DD NI/TA	0.21-0.24 0.26-0.28 0.40-0.42 0.42	H	No No No Yes
Mingo (1976)	NI/TA	0.006	H	No
Heggested and Mingo (1976)	Interest rate on new car loan Monthly service charge on demand deposits	0.17 0.194	H times a dummy variable for areas with low H	Yes Yes
Heggested (1977)	NI/TA	0.08	CR3	Yes
Whitehead (1977)	IL/TL IT/TS NI/C	0.39-0.45 0.37-0.45 0.39-0.43	N CR3 H	No Yes No
Whitehead (1978)	IL/TL IT/TS NI/C	0.160- 0.263 0.095- 0.129 -0.025- 0.066	CR3	Yes No No
Gradly and Kyle (1979)	SC/DD IT/TS IL/TL	0.34 0.25 0.37	H	No No No
Savage and Rhoades (1979)	NI/TA IL/TL SC/DD IT/TS	0.160 0.131 0.096 0.210	CR3	Yes Yes Yes
Rhoades and Rutz (1979)	IL/TL SC/DD NI/TA	O.22 0.19 0.05	CR3	Yes Yes No
Rose and Scott (1979)	IL/TL IT/TS	0.087 0.034	N CR1	No Yes
Hannan (1979a)	Wage and salary expenditure Number of banks employees	0.91-0.93	Dummy variable when CR3 is greater than 63%	Yes Yes
Hannan (1979b)	Interest rate paid on passbook saving a/cs	Used Tobit maximum likelihood analysis	CR3 H N	Yes No Yes
Glassman and Rhoades	NI/TA	0.12-0.13	CR3	Yes

(1980)				
Rhoades (1980)	NI/TA	0.02	CR3	Yes
Osborne and Wendel (1981)	SC/DD Service charge rates on DD	0.30	H	No
Rhodes (1981)	IL/TL SC/DD IT/TS NI/TA NI/C	0.14-0.32 0.08-0.23 0.12-0.24 0.05-0.18 0.10-0.29	CR3	Yes in 4 yrs Yes in 3 yrs Yes in 4 yrs No in all yrs Yes in 2 yrs
Rhodes and Savage (1981)	NI/TA	0.05	CR3 of deposits in the state in which bank is located	No
Rhoades and Rutz (1982)	NI/TA Coefficient of variation of NI/TA Equity / asset ratio Loan / asset ratio Net loan asset / total loans	0.003-0.06	CR3	Yes Yes Yes Yes No
Rhoades (1982a)	NI/TA	0.0034	CR3	Yes
Marlow (1982)	Interest rates on residential mortgage loans	0.25-0.31	N CR3 CR5	Yes Yes Yes
Kwast and Rose (1982)	NI/TA	0.42-0.58	H	Yes
Smirlock (1983)	NI/TA NI/C	0.03-0.05	CR3	No
Hannan (1984)	Passbook saving rate Total weekly banking hours	0.04-0.05	H	Yes No

NI = net income.
TA = initial asset.
C = capital.
IL = interest and fees on loans.
TL = total loans.
IT = interest payment on time and saving deposits
TS = time and saving deposits.
SC = revenue from service charge on demand deposits.
DD = demand deposits.
CR(n) = n-firms concentration ratio
H = Herfindahl index.
N = Number of loans in the market.

5.3.1.1 Profitability, concentration, and market share

To overcome the limitations of early studies Smirlock (1985) utilises profit rates and takes into account both market share and concentration. He suggests that market concentration is not a signal of collusive behaviour but rather of superiority of efficient leading firms in the industry. Accordingly, a positive coefficient of market share together with an insignificant coefficient for concentration indicates a support for the ES hypothesis. This is based on the notion that firms with high market share are more efficient and earn rents because of their efficiency, in contrast with the rest of firms. On the other hand, an insignificant coefficient for market share and a positive coefficient for concentration provide support for the structure performance hypothesis. In this case, rents reflected in higher profitability are the result of market concentration. The findings of Smirlock (1985) support the ES hypothesis reporting a positively significant market share variable and insignificant concentration variable. Other negative results for banking concentration are reported by Evanoff and Fortier (1988) who argue that the major linkage is between market share and profitability. Contrary to the above findings, Clark's (1986) results are highly supportive of the SCP model, with quantitatively a large and statistically significant estimated coefficient of market concentration.

5.3.1.2 Deposit Rates, concentration and market share

Despite the criticisms for having the interest rate as a measure of bank performance, a portion of more recent studies examine the SCP hypothesis in the retail deposit markets by utilising deposit rates interest [Berger and Hannan (1989, 1992), Calem and Carlino (1991), Jackson (1992)]. For the structure performance hypothesis to hold a negative price-concentration relationship must exist, as prices employed in these tests are paid *to* consumers (instead of being paid by consumers). In the case of the efficient hypothesis, if banks in concentrated markets are more efficient on average in gathering deposits and transforming them into profitable investments, then a positive relationship between price and concentration must exist. The findings by Berger and Hannan (1989) and Calem and Carlino (1991) are strongly consistent with the implications of the SCP hypothesis. However, according to Berger and Hannan, these price concentration results do not rule out the possibility that the efficient structure hypothesis also plays a role in the profit concentration relationship.

Jackson (1992) employes the same data set as Berger and Hannah (1989) and finds that the sign and the significance of the estimated coefficient on the market structure variable, changed for different subsets of data. In the low concentration group the coefficient is found to be negative, large and significant, while for the middle concentration group it is small and insignificant. The estimated coefficient for the high concentration group is positive and significant. This, in turn, suggests that the price concentration relationship is not linear over the relevant range, and may be U-shaped. He concludes that results can not support the SCP. On the other hand, results could support the ES as high level of market concentration may signal the gaining of market share by most efficient firms.

In their reply to Jackson (1992), Berger and Hannan (1992) utilise the same data set and find that, at least on average, the price concentration relationship for deposits is negative. Results suggest that the price concentration relationship varies over time and the argument that the relationship becomes positive at high levels of market concentration.

5.3.1.3 Direct test for the SCP, RMP, ESX, and ESS

All the aforementioned studies face a common limitation: the exclusion of direct measures of efficiency hypotheses in the performance equation. The exclusion of the efficiency variables generates several problems in estimating and interpreting the statistical results. First, no clear distinction can be made between various hypotheses. Some researchers argue that the common finding of a positive coefficient for market share and an insignificant coefficient for concentration provides justification for the acceptance of the relative market power hypothesis, which relates market share to market power. On the other hand, several authors consider this as an acceptance of either the ESX or the ESS hypothesis, because they relate the market power variable to the excluded efficiency variables. Further, a necessary condition for the efficiency hypotheses to hold is that efficiency is positively related to concentration and/or market share [Berger 1995a].

In an attempt to overcome this problem, some later studies include a direct measure of efficiency in the profitability equation. Allen and Hagin (1989) develop a relative direct cost measure of scale economies to examine the hypothesis that market share increases profits through the benefits of scale economies. The results provide minor support in favour for the hypothesis. Berger (1991) enters the debate between market power and ES explanations of the profit-concentration relationship in banking, by adding direct measures of X-efficiency to the empirical analysis. Results are consistent with the X-efficiency version of the ES hypothesis and limited support is also found for two of the other hypotheses, SCP and RMP. However, none of the hypotheses tested appear to be of overwhelming importance in explaining bank profits.

The major development in tests for the market structure and profit relationship emerges with the direct inclusion of structure and efficiency variables in the analysis. One of the first studies that includes both structure and efficiency variables in the tests is Berger and Hannan (1993). With the use of US data for the year 1985 they calculate direct efficiency measures in regressions of survey and profit data on market structure, they compare the results obtained from profit and survey data regressions, and analyse the effects of efficiency on market structure. From their results it is clear that the SCP is strongly supported by the data. In regressions explaining deposit and loan rates, the coefficients of concentration are both economically, and statistically significant and have the signs implied by the SCP hypothesis. Moreover, a negative relationship is found between X-efficiency and concentration, which supports the quiet life hypothesis, a supplement of the MP hypothesis. On the other hand, there is no support for the RMP hypothesis from the data. In most regressions the coefficient of the market share is either insignificant, or enters with the opposite sing to that predicted by the hypothesis.

As far as the efficiency measures are concerned, results are somewhat mixed. Both scale and X-efficiency measures are found to be positively associated with profitability. However, they are not associated with more favourable prices for consumers, as would be expected. Although this is not a necessary condition for the hypothesis to hold, it raises certain doubts about the ES hypothesis. Moreover, the data does not support a necessary condition for the hypothesis: the market structure variables are not consistently positively correlated with the efficiency measures.

A later study by Berger (1995a) uses two profitability measures, return on assets and return on equity. As a concentration measure the Herfindahl index is applied and the distribution-free approach is utilised for calculating X-efficiency and scale efficiency. Finally, a set of control variables is used to account for differences across markets. Tests of the MP and ES hypotheses are performed by regressing profits against measures of concentration, market share, X-efficiency, and scale efficiency. The author also examines whether efficiency affects market structure, a necessary condition for the ES hypotheses. The findings provide support for the two of the four hypotheses. There exists a partial support for the ESX; however, the test for the necessary conditions for this hypothesis to be valid is very weak. The other hypothesis supported by the results is the RMP. After controlling for the effects of concentration and efficiency the market share variable is positively related to profitability in most cases. From the results there is no evidence of support for the SPC or the ESS hypotheses. Contrary to the above results, Berger (1997) employs the same methodology and finds support for the SCP hypothesis and its 'quiet life addendum'.

Berger and Hannan (1998) examine whether market power exercise by banks in concentrated markets allows them to avoid cost minimisation, as opposed to the traditional concern about concentration in product markets, which focuses on the social loss associated with mispricing. Using data from the entire US banking industry for 1988, they regress efficiency on concentration and a set of control variables to estimate the aggregate cost from concentration-induced losses in bank efficiency. This cost is compared to estimates of social loss associated with misallocation of resources from non-competitive pricing (traditionally measured by the welfare triangle). The findings suggest that the estimated efficiency cost of concentration is several times larger than the social loss from mispricing.

5.3.2 International and European Studies

Most studies on market structure-performance relationship utilise data from the US market, and only a handful of papers employ data from other regions. The reason is the lack of publicly available regional banking market data. Therefore, as it is very difficult to obtain data on local banking markets, concentration ratios and individual market shares are calculated at a national level, as opposed to US banking studies.

A study that examines the relationship between market structure and profitability at international level is Bourke (1989). He employes data from 12 countries or territories,[67] and utilises a single form model on data from 90 banks, for the period 1972 to 1981. Based on previous literature he considers three different dependent variables, return on capital, return on assets, and a value-added return on total assets.[68] To account for concentration he estimates concentration ratio (CR3)[69], while the set of control variables employed consists of internal and external determinants of bank profitability (staff expenses, capital ratio, liquidity ratio, government ownership, interest rates market growth, and inflation). His results support the view that concentration is positively and moderately related to profitability. Molyneux and Thorton (1992) replicate Bourke's methodology to evaluate the determinants of European banks' profitability. Their sample covers data from 18 European banks for the period 1986 to 1989[70], and the results obtained are in agreement with the SCP hypothesis. The findings also confirmed Bourke's results, apart from the relationship between government ownership and profitability, where in contrast to Bourke's estimates they find positive relationship.

By utilisation of the same data set as in Molyneux and Thorton (1992), Molyneux and Forbes (1995) examine both the traditional SCP and the ES hypotheses. They estimate a profit equation with the use of return on assets, a ten-firm asset

[67] Countries and/or territories used: Australia, California, Massachusetts, New York, Canada, Germany, Ireland, England and Wales, Belgium, Holland, Norway, and Spain.
[68] Value added return on total assets: 1. the ratio of net income before taxes plus staff expenses to total assets, and 2. the ratio of net income before taxes plus staff expenses plus loan losses to total assets. This measure is used to capture the differences in accounting systems between countries.
[69] Concentration basis is taken as the share of the largest three banks of either total deposits or assets depending on data availability for each country.
[70] Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Liechtenstein, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweeden, Turkey, UK.

concentration ratio as a measure of market structure, and a firm specific share measure together with a vector of controlling variables to account for firm and market specific characteristics. The efficiency hypothesis is tested indirectly by the use of a market share variable, which should be greater than one, while the market structure variable should equal zero. Their results support the SCP hypothesis, with a significant and positive concentration ratio together with a negative and insignificant market share coefficient. Moreover, the capital asset ratio is always positive and significant, whereas the size variable is negative and statistically insignificant, pointing out that the size of a bank has no impact on profitability. Finally, they state that government owned banks are more profitable than privately owned ones.

Lastly, Goldberg and Rai (1996) employ data for the largest European banks established in eleven countries[71] for the years 1989 to 1992. They apply the same methodology as Berger (1995a). For measuring scale and X-efficiency they utilise the stochastic cost frontier. They use four measures of performance (return on assets and equity, and net interest margin on asset and equity), and two for concentration (three banks concentration, Herfindahl index). Finally, four control variables are employed. After dividing the sample between banks located in countries with high and low concentration, they conclude that there is evidence for the support X-efficiency hypothesis for banks located in low concentration countries. Finally, they state that results are very sensitive to the measure of performance used.

5.3.3 A summary of the empirical literature

It appears that results of empirical findings are mixed. Most early studies find a positive statistical relationship between market structure and measures of bank performance. However, due to the exclusion of a direct efficiency measure, their findings are interpreted differently. Further, each study includes different control variables. This, in turn, alters the statistical significance of the estimated regressions. What is clear from the literature is that little work is done outside the US market, and distinction between the four hypotheses is needed with the use of direct measures of both market structure and efficiency. In the current study the market structure-profit-

[71] UK, Germany, France, Swiss, Belgium, Finland, Austria, Italy, Sweden, Denmark, and Spain.

relationship is examined for the first time with the application of UK data, with direct measures for all the four hypotheses.

5.4 Methodology

5.4.1 Outline of the models describing the hypotheses

In this section we establish structural forms for efficient structure and the market power hypotheses to derive a single reduced-form model that includes all four hypotheses. Based on Berger (1995a) the structural model underlying the ESX and ESS is the following:

$$\pi_i = f_i(EFF_i, CV^1_{im}) + \varepsilon^1_{im} \tag{5.3}$$

$$MS_i = f_2(EFF_i, CV^2_{im}) + \varepsilon^2_{im} \tag{5.4}$$

$$CON_m = f_3\ (MS_i \text{ for all } i \text{ in } m) \tag{5.5}$$

where:

π = profitability per unit of output,

EFF = represents which ever efficiency concept is model,

MS_i = market share of bank i,

CON_m = concentration at market m,

CV_{im} = a set of control variables for i banks in market m, and

ε_{im} = error term.

Equation (5.3) suggests that profits differ due to variations created by efficiency. This, in turn, suggests that profit variations are generated by X-efficiency or scale efficiency, depending on whether the exact hypothesis is ESS or ESX. According to Berger (1995a) the positive relationship between profits and a measure of cost efficiency is a theory that could true or false, and should not be considered as an identity. Efficiency may not be related to profit if banks do not have systematic differences in costs or if these differences are offset by variations in revenues.

Equation (5.4) states that more efficient firms have higher market shares. This could occur for a number of reasons. First, if bank products within a local market are undifferentiated, each market may be in a competitive equilibrium with a common price equal to every bank's marginal cost. Hence, more efficient banks are larger and have greater market shares if they have lower marginal cost at every scale. Second, bank products may be differentiated by location, more efficient firms may set more favourable prices to consumers and, thus, attract customers from other locations. Lastly, larger efficiency of banks may result to a larger market share in equilibrium, due to past out-of-equilibrium behaviour, when more efficient firms gained share through price competition or through acquisition of less efficient banks.

Identity (5.5) indicates that the concentration measure is a (deterministic) transformation f_3 of market share into a concentration measure. This function is the same for all markets. Profitability and market structure are spuriously positively related because profit, concentration and market share are all positively related to efficiency. This can be seen from equations (5.3)-(5.5). Equation (5.3) states that more efficient firms have higher profitability; equation (5.4) asserts that more efficient firms have higher market share; where in identity (5.5) market share is positively related to concentration.

Concerning the market power hypotheses (MP), SCP and RMP, they are given by the following structural model:

$$\pi_i = f_4(Z_i, CV^4_{im}) + \varepsilon^3_{im} \qquad (5.6)$$

$$Z_i = f_5(MKTS_i, CV^5_{im}) + \varepsilon^4_{im} \qquad (5.7)$$

$$CON_m = f_3\ (MS_i \text{ for all } i \text{ in } m) \qquad (5.5)$$

where:

Z_i = is a vector of output prices, and

$MKTS$ = is a measure of market structure (concentration or market share, depending on the hypothesis tested).

The rest of the variables are defined as before.

In equation (5.6), as opposed to equation (5.3), profits are determined by differences in prices charged to consumers instead of cost differences. This does not rule out the possibility that efficiency affects profits, since a model of non-competitive behaviour does not preclude cost efficiency differences from affecting performance. Instead, the influence of efficiency on profits is viewed as less important than the exogenous effects of market power. Consequently, the EFF variable may appear as part of the set of control variables.

Equation (5.7) implies that prices charged to consumers are determined by market structure. Under the SCP hypothesis, the setting of prices is unfavourable to consumers under high degree of concentration, whereas under the RMP hypothesis, banks are able to exercise market power in pricing due to well-differentiated products brought about by advertising, location or other advantages; this results in a larger market share. Again, efficiency variables can affect market structure when included in the control variables, but their effect is assumed to be small.

The model of the MP hypothesis produces a spuriously positive relation between profit and market structure because concentration and market share are positively correlated by equation (5.6).

Most of the previous literature did not consider a direct measure of efficiency. Instead they estimated the following regression:

$$\pi_i = \alpha_0 + \alpha_1 CON_m + \alpha_2 MS_i + a_{3-n} CV_{im} + \varepsilon_{im} \tag{5.8}$$

In equation (5.8) the variable MS is considered as indicator for testing the RMP and the ES hypotheses. Concerning the RMP hypothesis, MS represents the market power of banks. When results generate a small and statistically insignificant coefficient of concentration, then profit and concentration are only spuriously related because both variables are correlated with the market share. Once the market share enters the equation, this relationship disappears. As far as the ES hypotheses are concerned, again they are tested through the MS variable based on the view that more efficient firms have greater market power. Support of the ES hypotheses comes from the notion that concentration and profitability are only related, because they are both correlated with efficiency. Once the MS (which in the current case is a representative of efficiency) enters the equation, this relationship disappears. To overcome this problem, the following equation is estimated, which includes a direct measure of all hypotheses, i.e.:

$$\pi_i = f_6(CON_m, MS_i, X-EFF_i, S-EFF_i, CV_{im}^6) + \varepsilon_{im}^7 \tag{5.9}$$

where:

π = profitability per unit of output,

$X\text{-}EFF_i$ = X-efficiency for firm i,

$S\text{-}EFF_i$ = scale efficiency for firm i

MS_i = market share of bank i,

CON_m = concentration at market m,

CV_{im} = a set of control variables for i banks in market m, and

ε_{im} = error term.

The reduced form equation (5.9) allows all four hypotheses to be valid simultaneously, where some of the explanatory variables become insignificant for some of the hypotheses. Under the ES hypothesis the coefficient of the appropriate efficiency variable is positive, whereas the coefficients of all other variables are either relatively small or zero. Under this hypothesis the zero coefficients of market share and concentration indicate that market structure has no direct effect on price and profit. Further, by utilising a model that directly incorporates efficiency, we assume that prices are set competitively and efficiency is strictly a function of lower cost with banks operating at efficient scale levels [Berger and Hannan (1993)]. Moreover, conditional on efficiency, market share and concentration are not related to profits. This is based on the notion that they are correlated with profits only because they reflect the effect of the control and efficiency hypotheses variables.

However, there are important limitations in the reduced form equation (5.9). It only tests one of the necessary conditions of the ES hypotheses. A positive relationship between efficiency and both profit and market structure variables must exist to explain the spurious profit structure relation. The following two equations are also tested to ensure that the necessary conditions hold. We estimate the following reduced form of concentration and market share from the ES model:

$$MS_i = f_7(X-EFF_i, S-EFF_i, CV_{im}^7) + \varepsilon_{im}^7 \tag{5.10}$$

$$CON_m = f_8(X-EFF_i, S-EFF, CV_{im}^8) + \varepsilon_{im}^8 \tag{5.11}$$

Under the ESS more efficient banks gain dominant market shares, as in (5.10). Further, following (5.11) efficient firms would be more often in concentrated markets since the large market shares create market concentration. The statistical significance of this positive relationship between market structure and efficiency will verify that more efficient firms have higher market share and cause concentration.

Under the MP hypotheses, market structure variables (concentration or market share) have a positive relationship with profits, whereas the rest of the coefficients will equal zero. When the RMP holds, the coefficient of concentration is zero because it is only related to profits through its correlation with the market share. When the SCP holds, concentration has a positive coefficient. The efficiency variables are appropriate exogenous variables, but are just viewed as relatively unimportant. Other critical differences from the ES model is the absence of a hypothesised positive causal relationship running from efficiency to market structure variables, as expressed by equations (5.10) and (5.11). According to Berger and Hannan (1993), and Berger (1995a, 1997) one can form the following 'reverse causation' model from market structure to efficiency, although it is not usually a part of the market power hypothesis:

$$X-EFF_i = f_9(CON_m, MS_i, CV_{im}^9) + \varepsilon_{im}^9 \tag{5.12}$$

$$S-EFF_i = f_{10}(CON_m, MS_i, CV_{im}^{10}) + \varepsilon_{im}^{10} \tag{5.13}$$

Equations (5.12) and (5.13) embody a version of Hicks' (1935) 'quiet life' hypothesis in which firms with greater market power may take part of the gains from non-competitive pricing not as profits, but as a more relaxed environment in which less effort is put on maximising cost efficiency. If the 'quiet life' life hypothesis holds, it tends to offset the positive profit-structure relationship, since gains from pricing are partially offset by cost increases from poorer efficiency ratios.

5.4.2 Description of the efficiency measures

A detailed analysis of the scale and X-efficiency measures is given in Chapter 3 and 4, respectively. In this section we briefly sketch efficiency estimation. **X-efficiency** provides a measure of how effectively banks use their inputs to produce a given level of output. As already mentioned, the most important difficulty in estimating X-efficiency is to isolate inefficiency differences from differences due to random error that temporarily alter costs. To distinguish between the two, in the current study we use the stochastic econometric frontier approach. The key to this approach is the two-part composed error term, where one part represents inefficiencies and the other part for statistical noise. These two components are separated by assuming that inefficiencies are drawn from a one-sided distribution (usually the half-normal) and random fluctuations are drawn from a symmetric distribution (usually the normal). The logic is that inefficiencies must have a truncated distribution, because they only increase costs above frontier levels, while random noise can either decrease or increase costs.

The cost equation estimated for each year is specified as follows:

$$\ln TC_i = \ln C(Y_i, w_i) + \ln u_i + \ln v_i \tag{5.14}$$

where:

w_i = price of inputs for the bank i,

y_i = outputs for the bank i,

TC = total cost for the bank i,

v_i = random error,

u_i = inefficiencies, and

$C(y, w)$ = is a cost function with output quantity vector y and input price vector w.

Equation 5.14 is estimated for each year by use of a translog specification. As in equation (5.14), u_i measures inefficiencies in the empirical test we substitute X-EFF defined in equation (5.9), that captures the X-efficiency level of each bank, with the X-INEFF inefficiency variable[see Goldberg and Rai (1996)]. The estimate X-INEFF, derived from the stochastic cost frontier, represents an inefficiency measure for each bank. As a result, the coefficient of X-INEFF in the regression of the efficient structure hypothesis will have the opposite sign from X-EFF specified in equations (5.9), (5.10), and (5.11). Hence, when the predicted sign of the coefficient of X-INEFF is negative, this implies that the lower the inefficiency the higher the profits. To test the validity of the results in different efficiency measures, we utilise two different measures both based on the stochastic frontier approach. The X-INEFF1 is based on the assumption that inefficiencies are half-normally distributed, whereas X-INEFF2 assumes that inefficiencies follow an exponential distribution.

Scale efficiency indicates whether banks with similar production and management technology operate at optimal economies of scale. Economies of scale are computed from the parameters of a system of equations consisting of the cost function (5.14) and the input share equation. The scale measure is given by the sum of the first-order partial derivatives of the estimated cost function, i.e.:

$$SE = \sum_{k=1}^{m} \frac{\partial \ln TC}{\partial \ln Y_k} \tag{5.15}$$

The aforementioned scale measure is estimated for each bank at its respective level of output, for each year. When SE < 1, banks operate below optimal scale level and have the ability to lower costs by increasing output further. When SE > 1, banks are required to downsize to achieve optimal input combinations. Because both SE >1 and SE <1 imply inefficiencies, a measure of inefficiency, S-INEFF, is used in the actual regression [see Goldberg and Rai (1996)], i.e.:

S-INEFF = SE - 1 when SE >1,

and

S-INEFF = 1-SE when SE < 1.

Again, the sign of the coefficient will be opposite to that predicted in equation (5.9), i.e. a negative sign. Consequently, the further the bank is from its efficient scale, the lower the profitability. This measure assumes that as banks deviate from the efficient scale the impact of efficiency is similar, i.e. banks' profitability increases monotonically whether they are downsizing or increasing in size towards the optimal scale [see Goldberg and Rai (1996)].

5.5 Data and Variables

The data employed is the same as the previous chapters (for a detail analysis on the data, see Chapter 2). To estimate the market concentration and the market share of each individual bank for each year, total indices are required, i.e. the total market deposits, collected from the annual financial statistics.

As far as the rest of the variables are concerned, a brief description is given below. To measure bank performance, two profitability ratios are used: the return on assets (ROA) and the return on equity (ROE). Return on assets is the ratio of net income divided by total assets and measures the efficiency of total assets employed by a firm. Return on equity is the ratio of net income to shareholders' equity, and appraises the efficiency of common shareholders' equity utilised by the firm.[72]

To estimate the market concentration, the concentration ratio is applied. The three-bank concentration ratio (CR3) is defined as the ratio of total deposits of the three largest banks to the total deposits of the UK banks. To measure the market share (MS) of each bank, we calculate the proportion of each individual bank's deposit in total market deposit.

For efficiency variables there are to separate measures. The first one concerns the X-efficiency (X-EFF) variable. As mentioned in the Chapter 4, X-efficiency indicates how effectively banks use their inputs to produce a given set of outputs, i.e. the ability of bank managers to maximise revenue or minimise cost. Due to the variations of cost frontiers from bank to bank, we need to estimate a stochastic cost function. Here the stochastic econometric frontier is applied, where the methodology and the data used are described in Chapter 4. The findings from the estimated cost frontier indicate the deviations from the efficient frontier. Accordingly, the X-EFF variable is replaced by X-INEFF. Consequently, the X-inefficiency variable will have a negative relationship with profit, i.e. increases in X-inefficiency will decrease profits.

Next, we need a measure for scale efficiency (S-EFF). This indicates whether banks with similar production and technologies operate at optimal economies of scale. Economies of scale imply the decline of cost when output increases. The methodology and the data used in estimating this variable is described in Chapter 3. When the estimate of S-EFF is greater than zero, banks have to lower production to achieve the optimum level of output. When S-EFF is lower than zero, banks have to increase production to achieve the optimum level of output. As this measure indicates inefficiencies, so S-EFF will be replaced with S-INEFF, which has a negative relationship with profitability.

Finally there is a set of three control variables. The first one is the size variable (LTA), represented by the logarithm of total assets of each individual bank to control for differences in size and to capture the ability of banks to diversify [Mester (1996)]. A positive relationship with profits exists if there are positive economies of scale, while there is a negative relationship if diversification leads to lower risk, and consequently, lower returns.

The second one is a risk measure (RISK), given by the ratio of total liabilities to total assets. A high risk ratio suggests less capital and higher leverage. This, is turn, results to high borrowing cost and hence to lower profitability. On the other hand, a lower risk ratio will result in higher profitability. Consequently, the coefficient of this variable can be either positive or negative.

The last control variable (WAGES) and accounts for the employees' cost. This is used to capture the effect of wages and in turn of operating cost, on profits. It is

[72]These are the most commonly used profitability measures in previous literature [see Molyneux *et*

expected to have a negative relationship with profits, especially in the case where the ES hypotheses are tested because higher costs represent lower efficiency levels. This variable is measured as the ratio of total wages and salaries to the number of employees.

5.6 Empirical estimation

The current study includes in the examination of the market performance hypothesis, direct measures of the ES hypothesis. For X-inefficiency two different variables are considered: the first variable is calculated based on the assumption of half-normal distribution (X-INEFF1) *Model I*, while the second is calculated based on the assumption of exponential distribution of X-inefficiency (X-INEFF2) *Model II*. Two measures are considered to test the robustness of the results to different estimates.

The empirical analysis focuses on the estimates of equations (5.9) to (5.13). Tables 5-3 and 5-4 provide the results from the estimation of the reduced form equation (5.9) using as a dependent variable ROA and ROE, respectively. However, before estimating (5.9), equation (5.8) is first estimated for both ROA and ROE to test the effect of the exclusion of efficiency variables on the regressions. Further, this replicates the current literature that finds MS to dominate the CONC and interprets this as a support for all hypotheses, except of the SCP.

The findings from these estimates are presented in Table 5-2. Tables 5-5 and 5-6 present the empirical results of estimating equations (5.10) and (5.11), while Table 5-7 displays the results of equations (5.12) and (5.13).

For each year we examine separate cross-section regressions and the efficiency and profitability measures to provide, first, a comprehensive treatment of the industry, and, second, to determine whether the results are stable over time and robust to measures selected. Similarly to many of the previous studies, the results are not unambiguous. As can be seen from Tables 5-3 and 5-5, ROA and ROE provide

---

al.(1996)]

different results for most of the years. For example, with ROE as the dependent variable, for the 1994 data, none of the hypotheses seem to be valid. However, when ROA is the dependent variable, the concentration variable is found to be positively statistically significant, and hence the SCP hypothesis is accepted. Results are also sensitive to the measure of X-efficiency used. Although some hypotheses are accepted when equation (5.9) is used, when the additional conditions are tested the results change. Last but not least the results are unstable over time. Details of the test are presented below. First, we consider the results from estimating equation (5.8), presented in Table 5-2.

Table 5-2: Regressions of ROA and ROE on Concentration, Market Share, and other control variables.

Year	Variable	CONC	MS	LTA	WAGES	RISK	R^2
1995	*ROA*	-0.123 (0.11)	-0.113 (0.19)	0.005 (0.01)	-0.1E-4 (0.2E-3)	0.007 (0.003)	0.15
	ROE	-0.846 (0.49)*	-0.497 (0.85)	0.042 (0.02)*	-0.005 (0.002)	0.063 (0.01)	0.44
1994	*ROA*	0.112 (0.027)*	-0.039 (0.04)	0.002 (0.001)*	-0.1E-3 (0.6E-4)	-0.098 (0.02)*	0.29
	ROE	-0.375 (0.62)	-0.397 (0.95)	0.078 (0.03)*	-0.002 (0.001)	-0.741 (0.53)	0.22
1993	*ROA*	0.061 (0.05)	-0.065 (0.08)	0.003 (0.002)	0.2E-4 (0.1E-3)	-0.091 (0.04)*	0.08
	ROE	-0.951 (0.44)*	-1.012 (0.72)	0.086 (0.02)*	0.3E-3 (0.001)	-0.551 (0.38)	0.23
1992	*ROA*	-0.021 (0.10)	0.013 (0.16)	0.6E-3 (0.01)	0.3E-4 (0.3E-3)	0.001 (0.01)	0.01
	ROE	-1.564 (1.10)	1.391 (8.39)	0.035 (0.25)	0.003 (0.01)	0.070 (0.33)	0.01
1991	*ROA*	0.068 (0.08)	0.068 (0.08)	0.002 (0.003)	-0.6E-4 (0.3E-3)	-0.088 (0.06)	0.03
	ROE	-0.061 (1.11)	-0.035 (0.21)	0.057 (0.05)	0.2E-3 (0.004)	-0.916 (0.91)	0.03
1990	*ROA*	-0.125 (0.06)*	-0.056 (0.10)*	0.002 (0.003)	-0.3E-4 (0.2E-3)	0.085 (0.06)	0.08
	ROE	-0.909 (0.44)*	-0.778 (0.68)	0.033 (0.02)	0.7E-3 (0.001)	0.334 (0.41)	0.09
1989	*ROA*	0.198 (0.07)*	-0.071 (0.09)	0.005 (0.004)	-0.1E-3 (0.2E-3)	-0.234 (0.07)*	0.20
	ROE	0.366 (0.59)	-1.181 (0.81)	0.071 (0.03)*	-0.001 (0.001)	-1.314 (0.59)*	0.11
1988	*ROA*	0.136 (0.02)*	0.087 (0.04)*	-0.004 (0.001)*	-0.1E-3 (0.9E-4)*	-0.044 (0.02)*	0.33
	ROE	0.195 (0.24)	0.717 (0.35)*	-0.025 (0.01)*	-0.001 (0.8E-3)*	0.353 (0.21)*	0.15
1987	*ROA*	0.194 (0.06)*	-0.056 (0.06)	0.003 (0.002)	-0.2E-3 (0.2E-3)	-0.201 (0.06)*	0.24
	ROE	1.135 (1.00)	-1.492 (1.03)	0.089 (0.03)*	-0.002 (0.003)	-2.233 (0.99)*	0.15
1986	*ROA*	0.128 (0.02)*	0.044 (0.02)*	-0.003 (0.001)*	-0.2E-3 (0.3E-3)	-0.054 (0.01)*	0.39
	ROE	0.393 (0.34)	0.185 (0.48)	0.024 (0.01)	-0.002 (0.001)	-0.538 (0.20)*	0.22
1985	*ROA*	0.079 (0.02)*	0.037 (0.02)*	-0.003 (0.9E-3)*	-0.1E-3 (0.7E-4)*	-0.016 (0.01)	0.37
	ROE	0.168 (0.13)	0.072 (0.17)	0.8E-3 (0.1E-2)	-0.1E-2 (0.5E-3)	-0.035 (0.07)	0.10

1984	*ROA*	0.129 (0.02)*	0.021 (0.02)	-0.002 (0.001)*	-0.1E-3 (0.7E-4)	-0.087 (0.03)*	0.41
	ROE	0.045 (0.35)	0.131 (0.37)	-0.016 (0.02)	-0.7E-3 (0.001)	0.342 (0.41)	0.02
1983	*ROA*	0.164 (0.02)*	0.015 (0.03)	-0.9E-3 (0.001)	-0.8E-4 (0.7E-4)	-0.114 (0.02)*	0.45
	ROE	0.623 (0.53)	-0.075 (0.55)	0.018 (0.02)	-0.5E-3 (0.002)	-0.681 (0.54)	0.04
1982	*ROA*	0.081 (0.01)*	0.044 (0.02)*	-0.004 (0.9E-3)*	-0.1E-3 (0.6E-4)*	-0.083 (0.10)	0.30
	ROE	0.287 (0.13)*	0.196 (0.19)	-0.007 (0.01)	-0.001 (0.6E-3)	0.007 (0.08)	0.08
1981	*ROA*	4.751 (2.47)*	4.286 (3.40)	-0.271 (0.14)*	0.013 (0.01)	-0.551 (1.32)	0.09
	ROE	2.598 (1.38)*	2.449 (1.89)	-1.498 (0.78)*	0.073 (0.06)	-2.845 (2.36)	0.09

Notes: Standard errors in the parenthesis/* denotes significance at 5%.

From Table 5-2, out of 30 regressions the CONC coefficients is 17 times significant, out of which 12 times positively significant, constituting a partial support for the SCP hypothesis, over the period. On the other hand, the MS coefficients are negative in thirteen cases, and only five times positive and statistically significant. The results here provide a partial support for the SCP hypothesis, and no evidence for the acceptance of RMP hypothesis.

Focusing next on cases where the market structure variables are positively significant, we observe that results are more stable mostly for the concentration variable during 1980s. With the use of ROA as the profit variable the CONC is positive statistically significant supporting the SCP hypothesis in nine cases. These findings reflect a setting of prices in more concentrated markets less favourable to consumers due to greater market power. This is the case for years 1981 to 1989, and 1994. Hence it is clear that throughout the 1980s the SCP hypothesis has been dominant in the banking industry. In the case of the ROE, SCP is accepted only for 1981 and 1982. This, in turn, suggests that the findings are subject to the choice of profit measurement.

Considering now the control variables, again their signs are somewhat mixed throughout the time period. LTA (size variable) is twelve times statistically significant suggesting that size has an effect on profit. However, as can be seen from

Table 5-2 coefficients signs are not fixed. By definition, a positive relationship with profits indicates positive economies of scale, while a negative relationship shows that diversification leads to lower risk, and consequently, lower returns. During the first decade, in most of the case the size variable has a negative coefficient, suggesting that the larger banks experience lower profits. Towards the end of the decade and until mid 1990s the coefficients become positive, indicating that banks started to benefit from their size, and earn profits. This maybe partly due to the changes in the business conditions that occurred during that period. Next, the control variable that accounts for the cost of the employees is found insignificant, whereas in three times when it is significant it has a negative sign, suggesting that higher employment cost result in lower profits. Last, a high risk (RISK) ratio suggests less capital and high leverage. This, in turn, results to high borrowing cost and hence to lower profitability. On the other hand, a lower risk ratio will result in higher profitability. Indeed, the coefficient is found to significantly negative in 11 times, and only once significantly positive, indicating that banks in the sample have high borrowing costs. The results reported in Table 5-2 are in line with the early literature that finds some evidence of significant association between the market structure and measures of bank performance with the direction of influence of the market structure as indicated by the SCP hypothesis [e.g. Rhoades and Rutz (1979), Glassman and Rhoades (1980), Rhoades and Rutz (1982) Rhoades (1982a), Marlow (1982)].

To have more definitive results next we considered the results from the reduced form profit equation for each year. Equation (5.9) incorporates the reduced form of all four hypotheses, and each coefficient gives the marginal effect of each hypothesis on profitability. The findings are presented in Tables 5-3 and 5-4.

Table 5-3: Regressions of ROA and ROE on Concentration, Market Share, Efficiency and other control variables - *Model I.*

Yr.	Var.	CONC	MS	X-INEFF1	S-INEFF	LTA	WAGE	RISK	R^2
1995	*ROA*	-0.092	-0.114	0.089	-0.079	0.003	0.3E-3	6E-3	0.19
		(0.13)	(0.19)	(0.07)	(0.09)	(0.01)	(6E-4)	(2E-3)*	
	ROE	-0.719	-0.503	0.724	-0.45	0.028	0.9E-3	0.058	0.50
		(0.54)	(0.82)	(0.31)*	(0.31)	(0.02)	(0.003)	(0.01)	
1994	*ROA*	0.129	-0.016	0.025	-0.041	0.001	0.2E-3	-0.089	0.35
		(0.03)*	(0.04)	(0.02)	(0.02)*	(0.001)	(2E-4)	(0.02)*	
	ROE	0.042	-0.556	-0.604	-0.122	0.083	-0.001	-0.933	0.25
		(0.79)	(0.98)	(0.45)	(0.56)	(0.03)*	(0.01)	(0.55)*	
1993	*ROA*	0.063	-0.065	-0.002	-0.003	0.004	0.4E-4	-0.092	0.08
		(0.05)	(0.08)	(0.03)	(0.02)	(0.003)	(1E-4)	(0.04)*	
	ROE	-0.917	-1.037	-0.029	-0.112	0.086	0.7E-3	-0.562	0.23
		(0.45)*	(0.73)	(0.24)	(0.21)	(0.03)*	(0.001)	(0.39)	
1992	*ROA*	-0.034	0.017	-0.017	0.030	0.7E-3	-0.3E-4	0.001	0.02
		(0.10)	(0.16)	(0.06)	(0.04)	(0.01)	(0.3E-3)	(0.01)	
	ROE	-1.902	2.460	1.222	0.769	0.007	0.001	0.082	0.01
		(5.23)	(8.69)	(3.15)	(1.98)	(0.26)	(0.01)	(0.33)	
1991	*ROA*	0.058	-0.006	0.013	-0.019	0.003	-0.1E-4	-0.078	0.05
		(0.08)	(0.02)	(0.04)	(0.02)	(0.003)	(0.3E-3)	(0.06)	
	ROE	0.925	-0.031	-1.130	-0.189	0.061	0.8E-3	-1.301	0.14
		(0.14)	(0.21)	(0.47)*	(0.25)	(0.05)	(0.003)	(0.89)	
1990	*ROA*	-0.141	-0.052	0.035	0.009	0.9E-3	-0.5E-4	0.099	0.09
		(0.07)*	(0.10)	(0.05)	(0.01)	(0.01)	(0.2E-3)	(0.06)	
	ROE	-0.889	-0.633	0.040	-0.127	0.034	0.001	0.335	0.15
		(0.47)*	(0.70)	(0.38)	(0.10)	(0.03)	(0.001)	(0.43)	
1989	*ROA*	0.195	-0.056	0.046	-0.008	0.004	-0.7E-4	-0.227	0.20
		(0.07)*	(0.10)	(0.08)	(0.02)	(0.003)	(0.2E-3)	(0.07)	
	ROE	0.418	-1.135	-0.037	-0.056	0.069	-0.9E-3	-1.291	0.11
		(0.62)	(0.84)	(0.71)	(0.14)	(0.03)*	(0.002)	(0.59)*	
1988	*ROA*	0.138	0.088	0.002	0.005	-0.004	-0.1E-3	-0.048	0.34
		(0.03)*	(0.03)*	(0.03)	(0.01)	(0.001)*	0.9E-4	(0.02)*	
	ROE	0.227	0.711	-0.062	0.056	-0.025	-0.001	0.31	0.17
		(0.24)	(0.35)*	(0.26)	(0.06)	(0.01)*	(0.9E-3)	(0.22)	
1987	*ROA*	0.194	-0.058	-0.003	0.001	0.003	-0.2E-3	-0.201	0.24
		(0.06)*	(0.06)	(0.01)	(0.01)	(0.002)	(0.2E-3)	(0.05)*	
	ROE	1.106	-1.495	-0.7E-3	-0.077	0.089	-0.001	-2.166	0.13
		(1.01)	(1.06)	(0.17)	(0.18)	(0.04)*	(0.002)	(1.02)*	
1986	*ROA*	0.121	0.043	0.003	0.006	-0.003	-0.2E-3	-0.052	0.40
		(0.03)*	(0.02)*	(0.01)	(0.01)	(0.001)*	(0.1E-3)*	(0.01)*	
	ROE	0.421	0.176	-0.035	-0.022	0.024	-0.001	-0.539	0.23
		(0.36)	(0.49)	(0.10)	(0.09)	(0.02)	(0.001)	(0.20)	
1985	*ROA*	0.081	0.038	0.004	-0.003	-0.003	-0.1E-3	-0.016	0.33
		(0.02)*	(0.02)*	(0.01)	(0.01)	(0.001)*	(0.7E-4)*	(0.01)	
	ROE	0.167	0.048	0.074	-0.017	0.4E-3	-0.001	-0.043	0.18
		(0.12)	(0.17)	(0.03)*	(0.03)	(0.01)	(0.5E-3)*	(0.07)	

1984	*ROA*	0.120 (0.02)*	0.017 (0.02)	-0.004 (0.01)	0.005 (0.01)	-0.002 (0.002)*	-0.1E-3 (0.8E-4)	-0.081 (0.02)*	0.42
	ROE	0.002 (0.384)	0.098 (0.37)	-0.087 (0.12)	0.019 (0.08)	-0.015 (0.01)	-0.9E-3 (0.001)	0.396 (0.43)	0.03
1983	*ROA*	0.163 (0.03)*	0.015 (0.03)	0.6E-3 (0.01)	0.3E-3 (0.01)	-0.9E-3 (0.001)	-0.8E-4 (0.7E-4)	-0.114 (0.02)*	0.44
	ROE	0.524 (0.57)	-0.044 (0.56)	0.101 (0.10)	0.007 (0.109)	0.019 (0.02)	-0.5E-3 (0.001)	-0.636 (0.56)	0.05
1982	*ROA*	0.083 (0.01)*	0.046 (0.02)*	0.004 (0.01)	-0.002 (0.01)	-0.004 (0.00)*	-0.1E-3 (0.7E-4)*	-0.009 (0.01)	0.31
	ROE	0.325 (0.15)*	0.199 (0.19)	-0.012 (0.11)	-0.036 (0.04)	-0.008 (0.01)	-0.96E-3 (0.6E-3)	0.2E-3 (0.08)	0.09
1981	*ROA*	3.360 (2.02)*	4.10 (3.40)	1.901 (1.80)	1.273 (0.88)	-0.231 (0.15)	0.009 (0.01)	-0.774 (1.32)	0.15
	ROE	1.877 (1.44)	2.354 (1.89)	1.032 (1.00)	0.68 (0.49)	-1.287 (0.81)	0.05 (0.06)	-0.405 (0.73)	0.14

Notes: Standard errors in the parenthesis/* denotes significance at 5%.

Table 5-4: Regressions of ROA and ROE on Concentration, Market Share, Efficiency and other control variables - *Model II.*

Yr.	Var.	CONC	MS	X-INEFF2	S-INEFF	LTA	WAGE	RISK	R^2
1995	*ROA*	-0.071 (0.13)	-0.118 (0.19)	0.010 (0.09)	-0.078 (0.10)	0.003 (0.01)	0.4E-3 (0.7E-3)	0.006 (0.003)*	0.17
	ROE	-0.606 (0.56)	-0.618 (0.86)	0.354 (0.41)	-0.526 (0.43)	0.028 (0.02)	0.002 (0.003)	0.058 (0.012)*	0.24
1994	*ROA*	0.142 (0.03)*	-0.026 (0.04)	0.008 (0.009)	-0.043 (0.02)*	0.001 (0.001)	0.2E-3 (0.2E-3)	-0.098 (0.02)*	0.34
	ROE	-0.317 (0.76)	-0.371 (0.99)	-0.007 (0.22)	-0.075 (0.56)	0.076 (0.03)*	-0.001 (0.01)	-0.738 (0.54)	0.22
1993	*ROA*	0.069 (0.05)	-0.070 (0.08)	-0.017 (0.03)	-0.003 (0.02)	0.004 (0.003)	0.4E-4 (0.2E-3)	-0.098 (0.05)	0.08
	ROE	-0.824 (0.47)	-1.103 (0.73)	-0.223 (0.27)	-0.113 (0.21)	0.090 (0.03)*	0.8E-3 (0.001)	-0.644 (0.40)	0.24
1992	*ROA*	0.056 (0.07)	-0.002 (0.11)	-0.249 (0.03)*	-0.054 (0.02)*	-0.5E-3 (0.003)	-0.5E-4 (0.2 E-3)	0.6E-4 (0.01)	0.53
	ROE	2.665 (3.76)	0.449 (6.06)	-13.112 (1.80)*	-2.479 (1.33)*	-0.020 (0.18)	-0.6E-3 (0.01)	0.016 (0.23)	0.51
1991	*ROA*	0.053 (0.08)	-0.005 (0.02)	0.028 (0.04)	-0.020 (0.02)	0.002 (0.004)	0.5E-5 (0.3E-3)	-0.072 (0.07)	0.06
	ROE	0.428 (1.13)	-0.121 (0.22)	-0.872 (0.43)*	-0.205 (0.26)	0.076 (0.05)	0.3E-3 (0.003)	-1.226 (0.93)	0.09
1990	*ROA*	-0.141 (0.07)*	-0.053 (0.10)	0.025 (0.03)	0.010 (0.01)	0.001 (0.004)	-0.5E-4 (0.1E-3)	0.099 (0.06)	0.09
	ROE	-0.871 (0.46)*	-0.65 (0.70)	-0.001 (0.22)	-0.126 (0.10)	0.036 (0.03)	0.001 (0.001)	0.320 (0.42)	0.12
1989	*ROA*	0.198 (0.07)*	-0.055 (0.10)	0.067 (0.12)	-0.008 (0.02)	0.004 (0.004)	-0.7E-4 (0.2E-3)	-0.228 (0.07)*	0.21
	ROE	0.413 (0.62)	-1.133 (0.85)	-0.023 (1.05)	-0.060 (0.14)	0.069 (0.03)*	-0.9E-3 (0.002)	-1.289 (0.60)*	0.11
1988	*ROA*	0.138 (0.03)*	0.088 (0.04)*	0.002 (0.02)	0.005 (0.01)	-0.004 (0.001)*	-0.1E-3 (0.9E-4)	-0.048 (0.02)*	0.34
	ROE	0.227 (0.24)	0.695 (0.35)*	-0.193 (0.24)	0.061 (0.06)	-0.025 (0.01)*	-0.001 (0.9E-3)*	0.334 (0.22)	0.18
1987	*ROA*	0.192 (0.06)*	-0.061 (0.06)	-0.006 (0.01)	0.8E-3 (0.01)	0.003 (0.002)	-0.2E-3 (0.2E-3)	-0.197 (0.06)*	0.25
	ROE	0.991 (0.99)	-1.689 (1.03)	-0.258 (0.15)	-0.122 (0.18)	0.089 (0.04)*	-0.002 (0.003)	-1.947 (0.99)*	0.18
1986	*ROA*	0.122 (0.03)*	0.040 (0.04)	-0.003 (0.01)	0.005 (0.01)	-0.003 (0.001)*	-0.2E-3 (0.1E-3)*	-0.051 (0.01)*	0.40
	ROE	0.409 (0.36)	0.179 (0.50)	-0.020 (0.10)	-0.016 (0.09)	0.024 (0.02)	-0.002 (0.002)	-0.537 (0.20)*	0.22
1985	*ROA*	0.081 (0.01)*	0.040 (0.02)*	0.5E-3 (0.01)	-0.003 (0.01)	-0.003 (0.001)*	-0.2E-3 (0.7E-4)*	-0.016 (0.01)	0.28
	ROE	0.190 (0.13)	0.089 (0.17)	0.056 (0.03)	-0.016 (0.03)	-0.001 (0.01)	-0.001 (0.001)	-0.037 (0.07)	0.15

1984	*ROA*	0.121 (0.03)*	0.019 (0.02)	0.47E-3 (0.01)	0.005 (0.01)	0.020 (0.09)	-0.1E-3 (0.8E-4)	-0.083 (0.03)*	0.42
	ROE	0.002 (0.39)	0.097 (0.38)	-0.059 (0.12)	-0.002 (0.001)*	-0.014 (0.02)	-0.8E-3 (0.001)	0.378 (0.43)	0.03
1983	*ROA*	0.163 (0.03)*	0.015 (0.03)	-0.3E-3 (0.01)	0.4E-3 (0.05)	-0.9E-3 (0.001)	-0.8E-4 (0.7E-4)	-0.114 (0.02)*	0.44
	ROE	0.575 (0.57)	-0.052 (0.56)	0.058 (0.09)	0.004 (0.11)	0.018 (0.02)	-0.5E-3 (0.002)	-0.655 (0.56)	0.04
1982	*ROA*	0.084 (0.02)*	0.045 (0.02)*	0.002 (0.01)	-0.003 (0.01)	-0.004 (0.001)*	-0.13E-3 (0.7E-4)*	-0.009 (0.01)	0.30
	ROE	0.326 (0.14)*	0.185 (0.19)	-0.078 (0.09)	-0.038 (0.04)	-0.007 (0.01)	-0.001 (0.6E-3)*	0.006 (0.09)	0.11
1981	*ROA*	3.780 (2.10)*	4.582 (3.44)	1.802 (1.49)	1.308 (0.88)	-0.255 (0.14)*	0.009 (0.011)	-0.776 (1.31)	0.15
	ROE	2.111 (1.45)	2.627 (1.91)	1.006 (0.83)	0.700 (0.49)	-1.42 (0.82)*	0.049 (0.06)	-0.410 (0.73)	0.15

Notes: Standard errors in the parenthesis/* denotes significance at 5%.

When all variables are included in the profit equation, from Table 5-3 and 5-4, we observe that the CONC has forty-two positive coefficients with twenty-four of them statistically significant. On the other hand, the MS coefficients are twenty-nine times positive from of which nine are statistically significant. With regards to the ES hypothesis out of sixty cases, X-INEFF is thirty-one times positive, statistically significant only in five cases, while S-INEFF is twenty-four times positive, but again only in five cases significant. On average the CONC and MS coefficients do not change in a meaningful way when the ES variables are added to the regressions. Furthermore the preceding results suggest that the MS coefficients in the commonly specified equation without the efficiency variables do not reflect the effects of efficiency.

From the findings there is some uncertainty as to the quantitative and qualitative importance of the inclusion of the efficiency variables. Contrary with the results of Berger (1995a, 1997) but consistent with Berger and Hannan (1993), the inclusion of the efficiency variables does not alter significantly the R^2 of the regressions, where in most case increases by a very small percentage or remains the same. In any case, the small R^2s are persistent in the profit-structure relation literature [see Table 5-1]. Apparently this suggest that the variation in profits is due to other factors as well, such as portfolio choices and regulatory changes and not only in the structure of the industry.

Furthermore, in most of the years the efficiency variables are found insignificant or with the wrong sign. Consistently with the results for European studies [Molyneux and Thorton (1992) and Molyneux and Forbes (1995)] it is clear that in most years the dominant hypothesis is the traditional SCP. This suggests that the structure of the European banking system is different from the US, since in Europe the market tends to be more concentrated, dominated by large multi-branch universal banks. The findings further indicate that banks earn higher profits due to competitive imperfections in the market resulting from the oligopolistic structure of the industry.

Focusing now in each year separately, starting with 1981, when the ROA is considered as the dependent variable there exists a support for the SCP hypothesis, with the CONC variable positive and statistically significant for all X-efficiency measures. In both cases, all the other variables are found to be statistically insignificant with an exception of the size variable in *Model II* which becomes significant. Comparing these results with previous estimates (Table 5-2) we can see that they are consistent, i.e. they still support the SCP hypothesis, but there is 5% increase in the R^2. Proceeding with the estimates from the ROE, when the direct measures of the ES hypotheses are included in the equations, the CONC becomes statistically insignificant.

For 1982 to 1984, we have the same results. With the use of ROA both MS and CONC variables are positive and statistically significant, whereas the efficiency variables together with RISK are insignificant. The findings of a positive profit-concentration and profit-market share relationship when other factors affecting profits are controlled for, supports the SCP and RMP hypotheses, respectively. Accordingly, the findings for this period advocate that higher market share and concentration result in higher profitability. Both LTA and WAGES have a negative sign suggesting that a decrease in wage costs and bank size increases profits. Now, in the case of ROE, the RMP hypothesis is rejected, whereas the CONC is still significant together with WAGES for the second model.

Considering *Model I* for 1985 and 1986, we can see from Table 5-3 that in the case of ROA, both the two MP hypothesis are accepted in contract to the rejection of the ES hypothesis, with control variables being negative and significant. As regards the ROE, none of the variables seem to be significant in explaining profits for the current period. Again, consistent with previous studies, it is clear that the results are sensitive to the measure of profit utilised, as they produce totally different estimates [Molyneux *et al.* (1996), Berger and Hannan (1993), Berger (1997)]. Consistently in 1987, with ROA as dependent variable the data again supports the SCP hypothesis, whereas with ROE all four hypotheses are rejected.

For 1988 once more there is a support for both MS hypotheses with ROA, while with ROE only RMP is supported. In the first case profits are positively related to market share as well as concentration, whereas in the second case profits are only driven by market share. For the next year findings suggest that size and risk only explain ROE, with size having a positive relationship with profits and a negative one with risk. On the other hand, the SCP hypothesis is supported, when ROA is employed as the profit measure.

It is evident from the above that during the 1980s the dominant hypothesis is the SCP, suggesting that higher profits in the industry results from its concentrated structure. Nevertheless, following the changes in the UK banking during the 1980s, the empirical results for the first part of the 1990s provide a different picture. In 1990, although the CONC variable is found significant, it has a negative sign. The findings of a mostly negative profit-concentration relationship run strongly contrary to the SCP hypothesis. This in turn suggests that higher concentration leads to lower profitability. In addition, none of the control variables are found to be statistically significant.

In 1991, with ROE as dependent variable, the X-INEFF variable is found negatively significant for the first time indicating that the lower the level of inefficiency the higher the level of profit. Consistent with ESX hypothesis, this suggests that banks earn excess profits due to superior management or production technologies. When ROA is considered none of the variables are found to be significant. In 1992 *Model I* supporta the notion that none of the variables explains the profit variation. However, in *Model II* both ES hypotheses are accepted since X-INEFF and S-INEFF are found to be negative and statistically significant. As we noted earlier, the ES explanation for the positive relationship between profits, and concentration or market share, states that efficient firms increase in size and market share because of their ability to generate higher profits which usually leads to higher market concentration. The results suggest that profits increase due to superior management and technologies, and because of the more efficient supply of products, which is turn lower costs.

Data for 1993 support once more the SCP hypothesis, in both models, with the ROE as dependent variable, whereas in the case of ROA all hypotheses are rejected. In 1994 in the case of ROA both models support the ESS hypothesis. This states that banks have essentially equally good management and technology but some of them supply products at more efficient scale than others do. Hence they have lower unit cost and higher market profits. Proceeding with ROE, none of the hypotheses is accepted, and the only variable that explains profits is RISK. Finally, for 1995 and for both profit measures, none of the hypotheses is accepted.

In general the above findings are partially in line with the current literature. Most studies do not include direct measures for the ES hypothesess and find mixed results; Smirlock (1985) and Evanoff and Fortier (1988) find strong support of the ES hypothesis, while Clark (1986) Berger and Hannan (1989,1992), and Molyneux and Forbes (1995) find evidence supporting the SCP hypothesis. In our case, during the 1980s in most cases the evidence support the validity of the SCP hypothesis: the positive relationship between profit and concentration reflects a price setting that is less favourable to consumers in more concentrated markets due competitive imperfections in those markets.

Considering the few studies that include all four hypotheses in their estimation, the current results are in line with Berger and Hannan (1993). Nevertheless, in some isolated cases during 1990s, there are some evidences in favour of the ES hypotheses. A possible explanation for these findings is that the structure of the UK banking industry started to change following the national and international deregulation pattern during the 1980s. Banks have faced new competitors and new forms of competition have developed. This intensifying competition increases the pressure for improvements in cost efficiency. As a result in an attempt to become competitively viable banks nowadays operate more efficiently.

As a next step, we run further regressions to examine whether market power or efficiency effects associated with MS that eliminates the positive CONC coefficients. According to Berger (1995a, 1997) this is the key point in the debate between MP and ES advocates in the literature. Tables 5-5 and 5-6 show the results for equations (5.10) and (5.11), which set the necessary conditions for the efficient structure hypothesis. Estimates of the market structure, MS, and market concentration, CONC, are regressed on X-INEFF, S-INEFF and other control variables. As mentioned in the previous section, under the efficient structure hypothesis, the relationship between inefficiency and concentration or market share should be negative.

Table 5-5: Regressions of Market Share and Market Concentration on X-inefficiency, Scale inefficiency and other control variables- *Model I.*

Year	Variable	X-INEFF1	S-INEFF	LTA	WAGES	RISK
1995	*CONC*	0.096 (0.09)	0.469 (0.09)*	0.037 (0.001)*	0.006 (0.003)*	-0.003 (0.7E-3)*
	MC	-0.036 (0.06)	-0.171 (0.06)*	0.003 (0.001)*	0.002 (0.002)	0.6E-3 (0.4E-3)
1994	*CONC*	0.225 (0.07)*	0.371 (0.08)*	0.005 (0.004)	0.460 (0.07)*	-0.003 (0.7E-3)*
	MC	-0.112 (0.06)*	0.026 (0.07)	0.021 (0.003)*	-0.6E-3 (0.7E-3)	-0.275 (0.055)*
1993	*CONC*	0.128 (0.07)*	0.043 (0.07)	0.008 (0.01)	0.2E-4 (0.4E-3)	0.520 (0.09)*
	MC	-0.038 (0.05)	-0.025 (0.04)	0.021 (0.004)*	-0.2E-3 (0.3E-3)	-0.286 (0.06)*
1992	*CONC*	0.052 (0.10)	0.098 (0.06)	0.041 (0.002)*	0.1E-3 0.4E-3	0.002 (0.01)
	MC	-0.064 (0.05)	-0.007 (0.03)	0.020 (0.003)*	-0.3E-3 (0.2E-3)	-0.006 (0.01)
1991	*CONC*	0.147 (0.05)*	0.002 (0.03)	0.011 (0.005)*	0.001 (0.4E-3)*	0.505 (0.08)*
	MC	0.111 (0.06)*	-0.30 (0.16)*	0.075 (0.03)*	-0.002 (0.002)	-0.930 (0.44)*
1990	*CONC*	0.326 (0.10)*	-0.002 (0.03)	0.006 (0.01)	0.5E-3 (0.4E-3)	0.650 (0.09)*
	MC	-0.177 (0.07)*	0.025 (0.02)	0.024 (0.004)*	-0.6E-3 (0.2E-3)*	-0.301 (0.06)*
1989	*CONC*	0.337 (0.15)*	0.032 (0.03)	0.006 (0.01)	0.4E-3 (0.4E-3)	0.693 (0.09)*
	MC	-0.202 (0.11)*	0.014 (0.022)	0.025 (0.004)*	-0.7E-3 (0.3E-3)*	-0.316 (0.07)*
1988	*CONC*	0.192 (0.155)	-0.035 (0.03)	0.013 (0.005)*	0.9E-3 (0.5E-3)*	0.638 (0.10)*
	MC	-0.109 (0.11)	0.010 (0.03)	0.022 (0.004)*	-0.8E-3 (0.3E-3)*	-0.277 (0.07)*
1987	*CONC*	-0.8E-3 (0.02)	-0.012 (0.02)	-0.5E-3 (0.004)	-0.2E-3 (0.4E-3)	0.871 (0.07)*
	MC	-0.011 (0.02)	-0.7E-3 (0.02)	0.027 (0.004)*	-0.4E-3 (0.3E-3)	-0.369 (0.06)*
1986	*CONC*	0.048 (0.04)	0.091 (0.03)*	0.026 (0.004)*	-0.5E-4 (0.6E-3)	0.399 (0.07)*
	MC	-0.037 (0.03)	-0.025 (0.03)	0.014 (0.003)*	-0.7E-3 (0.7E-3)	-0.138 (0.05)

1985	*CONC*	-0.007 (0.04)	0.032 (0.04)	0.030 (0.004)*	0.9E-4 (0.7E-3)	0.426 (0.08)*
	MC	0.022 (0.03)	0.014 (0.03)	0.014 (0.003)*	-0.70E-3 (0.5E-3)	-0.173 (0.06)*
1984	*CONC*	-0.010 (0.04)	0.082 (0.03)	0.2E-4 (0.004)	-0.8E-3 (0.5E-3)	0.990 (0.07)*
	MC	-0.024 (0.04)	0.013 (0.03)	0.034 (0.01)	-0.5E-3 (0.5E-3)	-0.463 (0.07)*
1983	*CONC*	0.023 (0.02)	0.067 (0.02)*	0.8E-3 (0.004)	0.1E-5 (0.4E-3)	0.800 (0.06)*
	MC	-0.010 (0.02)	-0.030 (0.02)	0.032 (0.004)	-0.3E-3 (0.4E-3)	-0.414 (0.07)*
1982	*CONC*	0.144 (0.12)*	0.101 (0.04)*	0.027 (0.005)*	0.8E-3 (0.6E-3)	0.384 (0.08)*
	MC	-0.117 (0.09)	-0.031 (0.03)	0.018 (0.004)	-0.8E-3 (0.5E-3)	-0.180 (0.06)
1981	*CONC*	0.078 (0.11)	0.131 (0.05)*	0.034 (0.005)*	0.6E-3 (0.7E-3)	0.310 (0.07)*
	MC	-0.087 (0.08)	-0.026 (0.04)	0.015 (0.004)*	-0.8E-3 (0.5E-3)	-0.137 (0.05)

Notes: Standard errors in the parenthesis/* denotes significance at 5%.

Table 5-6: Regressions of Market Share and Market Concentration on X-inefficiency, Scale inefficiency and other control variables- *Model II.*

Year	Variable	X-INEFF2	S-INEFF	LTA	WAGES	RISK
1995	*CONC*	0.016 (0.11)	0.478 (0.10)*	0.038 (0.002)*	-0.003 (0.7E-3)*	0.006 (0.003)*
	MC	0.055 (0.07)	-0.192 (0.06)*	0.002 (0.001)	0.7E-3 (0.5E-3)	0.001 (0.002)
1994	*CONC*	0.023 (0.04)	0.406 (0.09)*	0.008 (0.004)*	-0.003 (0.8E-3)	0.466 (0.07)*
	MC	0.006 (0.03)	0.008 (0.07)	0.020 (0.003)*	-0.4E-3 (0.6E-3)	-0.29 (0.06)*
1993	*CONC*	0.210 (0.08)*	0.051 (0.06)	0.006 (0.01)	-0.4E-4 (0.4E-3)	0.546 (0.09)*
	MC	-0.044 (0.05)	-0.016 (0.04)	0.024 (0.003)*	-0.2E-3 (0.3E-3)	-0.183 (0.07)*
1992	*CONC*	0.132 (0.08)*	0.090 (0.06)	0.040 (0.002)*	0.1E-3 (0.4E-3)	0.002 (0.01)
	MC	-0.056 (0.03)*	-0.052 (0.04)	0.006 (0.001)*	-0.3E-3 (0.3E-3)	-0.006 (0.01)
1991	*CONC*	0.133 (0.06)*	-0.9E-3 (0.03)	0.010 (0.005)*	0.001 (0.4E-3)*	0.518 (0.09)*
	MC	-0.304 (0.32)	-0.262 (0.16)*	0.084 (0.03)*	-0.001 (0.002)	-0.965 (0.44)*
1990	*CONC*	0.161 (0.06)*	0.001 (0.03)	0.008 (0.01)	0.4E-3 (0.4E-3)	0.659 (0.09)
	MC	-0.085 (0.04)*	0.023 (0.02)	0.022 (0.004)*	-0.6E-3 (0.3E-3)*	-0.310 (0.06)
1989	*CONC*	0.409 (0.23)*	0.031 (0.03)	0.006 (0.006)	0.4E-3 (0.4E-3)	0.703 (0.09)*
	MC	-0.283 (0.17)	0.014 (0.02)	0.025 (0.004)*	-0.7E-3 (0.3E-3)	-0.350 (0.07)*
1988	*CONC*	0.086 (0.15)	-0.036 (0.04)	0.014 (0.005)*	0.9E-3 (0.5E-3)*	0.653 (0.10)*
	MC	-0.078 (0.10)	0.011 (0.03)	0.021 (0.00)*	-0.8E-3 (0.3E-3)	-0.279 (0.07)*
1987	*CONC*	-0.007 (0.02)	-0.013 (0.02)	-0.6E-3 (0.004)	-0.2E-3 (0.4E-3)	0.876 (0.0)*
	MC	-0.014 (0.02)	-0.002 (0.03)	0.027 (0.004)*	-0.4E-3 0.3E-3	-0.363 (0.06)*
1986	*CONC*	0.022 (0.05)	0.084 (0.04)*	0.028 (0.01)*	-0.8E-4 (0.6E-3)	0.405 (0.07)
	MC	-0.036 (0.03)	-0.021 (0.03)	0.013 (0.003)*	-0.7E-3 (0.5E-3)	-0.135 (0.05)

1985	*CONC*	-0.018 (0.04)	0.032 (0.04)	0.031 (0.005)*	0.1E-3 (0.7E-3)	0.426 (0.08)*
	MC	0.005 (0.03)	0.015 (0.03)	0.014 (0.003)*	-0.7E-3 (0.5E-3)	-0.170 (0.06)*
1984	*CONC*	-0.013 (0.04)	0.083 (0.03)*	0.1E-3 (0.004)	-0.8E-3 (0.5E-3)*	0.990 (0.07)*
	MC	-0.036 (0.04)	0.013 (0.02)	0.034 (0.01)*	-0.5E-3 (0.5E-3)	-0.464 (0.07)*
1983	*CONC*	0.008 (0.02)	0.068 (0.02)*	0.6E-3 (0.004)	-0.2E-5 (0.4E-3)	0.804 (0.06)*
	MC	-0.006 (0.02)	-0.030 (0.02)	0.032 (0.004)	-0.3E-3 (0.4E-3)	-0.417 (0.07)*
1982	*CONC*	0.057 (0.08)	0.106 (0.04)*	0.028 (0.01)*	0.8E-3 (0.6E-3)	0.398 (0.08)*
	MC	-0.067 (0.07)	-0.035 (0.03)	0.018 (0.004)*	-0.8E-3 (0.5E-3)	-0.189 (0.06)*
1981	*CONC*	0.008 (0.09)	0.134 (0.05)*	0.035 (0.005)*	0.6E-3 (0.7E-3)	0.321 (0.07)*
	MC	-0.088 (0.07)	-0.030 (0.04)	0.015 (0.004)*	-0.8E-3 (0.5E-3)	-0.139 (0.05)*

Notes: Standard errors in the parenthesis/* denotes significance at 5%.

With the concentration as the dependent variable, the X-INEFF is found in twenty four times out of thirty to be positive, with eleven cases being statistically significant. However, in all cases the coefficients are positive. When the dependent variable is MS in only five cases the coefficient is positive with one of them being significant. The rest have a negative sign and five times it is found significant. When the CONC is the dependent variable the S-INEFF in twenty-five case is positive, with thirteen of them being statistically significant. In the case of MS twelve coefficients are positive but none significant and eighteen negative, as the relationship requires, with two of them statistically significant. This does not support earlier results on the ESX and ESS hypotheses. Although some of the coefficients have the correct sign, they are insignificant. Thus, overall, there is very little support for the ES hypothesis. Additionally, coefficients on CONC or MS are not substantially altered when efficiency variables are added in the regressions, suggesting that *is* the market power effects captured by MS drive the profit-concentration relationship, *not* efficiency [Berger (1995a)].

Table 5-6 shows the results for equation (5.12) and (5.13), which test Hicks' 'quiet life' hypothesis. Hicks (1935, p.8) suggests that *'the best of all monopoly profits is a quiet life'*. According to Berger and Hannan (1993), if this hypothesis holds, it tends to offset the profit structure relationship since gains from pricing are partially offset by cost increases from the poorer efficiency ratios. The hypothesis states that when firms enjoy greater market power and concentration, inefficiency follows not because of non-competitive pricing but because of a relaxed environment that produces no incentive to minimise costs. The dependent variables X-INEFF1, X-INEFF2, and S-INEFF are regressed against CONC, MS and other control variables. To recall, X-INEFF1 and X-INEFF2 measure the X-inefficiencies, and S-INEFF measures the scale inefficiencies. For the hypothesis to hold the signs of CONS and MS must be positive. Determination of which of the inequalities hold, (5.12) or (5.13), depends on the market structure variable associated with market power (concentration or market share), and on the dimension of efficiency at which banks with market power become unrestrained in pursuing.

Table 5-7: Regressions of X-inefficiency, Scale inefficiency on Market Share, Concentration and other control variables

Year	Variable	CONC	MS	LTA	WAGES	RISK
1995	*X-INEFF1*	0.232 (0.21)	-0.033 (0.36)	0.003 (0.01)	0.7E-3 (0.5E-3)	0.001 (0.01)
	X-INEFF2	0.295 (0.17)*	0.241 (0.29)	0.002 (0.01)	-0.2E-3 (0.4E-3)	-0.3E-3 (0.004)
	S-INEFF	0.656 (0.16)*	-0.067 (0.28)	-0.025 (0.01)*	0.006 (0.4E-3)*	-0.010 (0.003)*
1994	*X-INEFF1*	0.539 (0.19)*	-0.331 (0.29)	0.014 (0.008)*	-0.3E-3 (0.4E-3)	-0.324 (0.16)*
	X-INEFF2	0.241 (0.39)	0.207 (0.59)	-0.002 (0.017)	-0.3E-3 (0.9E-3)	0.090 (0.33)
	S-INEFF	0.742 (0.15)*	0.339 (0.23)	-0.026 (0.07)*	0.008 (0.3E-3)*	0.026 (0.13)
1993	*X-INEFF1*	0.428 (0.25)*	-0.105 (0.42)	0.006 (0.01)	0.4E-3 (0.6E-3)	-0.124 (0.23)
	X-INEFF2	0.468 (0.22)*	-0.307 (0.36)	0.016 (0.01)	0.2E-3 (0.5E-3)	-0.386 (0.187)
	S-INEFF	0.192 (0.29)	-0.198 (0.47)	0.004 (0.02)	0.004 (0.7E-3)*	-0.062 (0.25)
1992	*X-INEFF1*	0.016 (0.24)	-0.590 (0.39)	0.019 (0.01)	0.1E-3 (0.6E-3)	-0.005 (0.02)
	X-INEFF2	0.408 (0.27)	-0.158 (0.46)	-0.003 (0.01)	0.1E-3 (0.7E-3)	-0.006 (0.02)
	S-INEFF	0.415 (0.37)	-0.455 (0.61)	0.007 (0.02)	0.002 (0.001)*	-0.008 (0.02)
1991	*X-INEFF1*	0.859 (0.30)*	0.058 (0.03)*	0.002 (0.01)	0.1E-3 (0.9E-3)	-0.380 (0.25)
	X-INEFF2	0.555 (0.277)*	-0.050 (0.05)	0.019 (0.01)	-0.5E-3 (0.9E-3)	-0.411 (0.23)*
	S-INEFF	0.028 (0.58)	-0.207 (0.11)*	0.015 (0.03)	0.002 (0.002)	0.237 0.476
1990	*X-INEFF1*	0.391 (0.16)*	-0.363 (0.24)	0.019 (0.01)*	-0.1E-3 (0.4E-3)	-0.362 (0.15)*
	X-INEFF2	0.544 (0.28)*	-0.484 (0.42)	0.021 (0.02)	0.1E-3 (0.8E-3)	-0.547 (0.25)
	S-INEFF	0.293 (0.58)	1.026 (0.88)	0.015 (0.03)	0.003 (0.002)	-0.108 (0.53)
1989	*X-INEFF1*	0.189 (0.10)*	-0.183 (0.16)	0.007 (0.01)	-0.8E-4 (0.4E-3)	-0.083 (0.11)
	X-INEFF2	0.099 (0.08)	-0.124 (0.11)	0.005 (0.004)	-0.9E-4 (0.3E-3)	-0.041 (0.08)
	S-INEFF	0.756 (0.59)	0.847 (0.82)	-0.043 (0.03)	0.004 (0.002)*	0.427 (0.59)

1988	*X-INEFF1*	0.108 (0.12)	-0.104 (0.18)	0.002 (0.01)	-0.2E-4 (0.4E-3)	0.068 (0.11)
	X-INEFF2	0.028 (0.14)	-0.118 (0.20)	0.002 (0.01)	-0.3E-3 (0.5 E-3)	0.161 (0.12)
	S-INEFF	-0.436 (0.52)	-0.019 (0.76)	-0.003 (0.02)	0.003 (0.002)	0.841 (0.46)
1987	*X-INEFF1*	-0.036 (0.7851)	-0.386 (0.82)	0.013 (0.03)	0.162 (0.78)	-0.7E-3 (0.002)
	X-INEFF2	-0.036 (0.78)	-0.386 (0.82)	0.013 (0.03)	-0.7E-3 (0.002)	0.161 (0.78)
	S-INEFF	-0.381 (0.92)	-0.745 (0.96)	0.003 (0.04)	-0.003 (0.002)	0.696 (0.91)
1986	*X-INEFF1*	0.035 (0.49)	-0.571 (0.71)	0.010 (0.03)	-0.002 (0.002)	0.182 (0.29)
	X-INEFF2	-0.183 (0.45)	-0.681 (0.64)	0.013 (0.02)	-0.001 (0.002)	0.342 (0.26)
	S-INEFF	1.215 (0.55)*	0.500 (0.77)	-0.026 (0.03)	0.004 (0.002)*	-0.339 (0.32)
1985	*X-INEFF1*	0.187 (0.51)	0.536 (0.69)	-0.002 (0.03)	0.1E-3 (0.002)	0.162 (0.29)
	X-INEFF2	-0.171 (0.50)	-0.034 (0.68)	0.022 (0.03)	(0.5E-5) (0.002)	0.110 (0.27)
	S-INEFF	0.774 (0.56)	0.929 (0.75)	-0.039 (0.03)	0.6E-3 (0.002)	0.267 (0.32)
1984	*X-INEFF1*	-0.121 (0.40)	-0.243 (0.41)	0.002 (0.02)	-0.6E-3 (0.001)	0.400 (0.46)
	X-INEFF2	-0.167 (0.41)	-0.381 (0.43)	0.016 (0.02)	0.3E-4 (0.001)	0.227 (0.48)
	S-INEFF	1.663 (0.55)*	0.610 (0.58)	-0.032 (0.03)	0.006 (0.002)*	-0.976 (0.65)
1983	*X-INEFF1*	0.854 (0.70)	-0.261 (0.73)	-0.003 (0.03)	-0.1E-3 (0.002)	-0.371 (0.72)
	X-INEFF2	0.706 (0.83)	-0.345 (0.87)	0.006 (0.04)	0.1E-3 (0.002)	-0.377 (0.85)
	S-INEFF	1.800 (0.66)*	-0.598 (0.69)	0.003 (0.03)	0.7E-5 (0.002)	-1.118 (0.68)
1982	*X-INEFF1*	0.123 (0.17)	-0.179 (0.23)	0.004 (0.01)	-0.3E-3 (0.7E-3)	0.067 (0.10)
	X-INEFF2	0.022 (0.21)	-0.208 (0.29)	0.010 (0.01)	-0.6E-3 (0.9E-3)	0.082 (0.13)
	S-INEFF	1.013 (0.46)*	0.148 (0.63)	-0.020 (0.03)	0.001 (0.002)	-0.207 (0.28)
1981	*X-INEFF1*	0.061 (0.18)	-0.179 (0.25)	0.006 (0.01)	-0.2E-3 (0.8E-3)	0.105 (0.10)
	X-INEFF2	-0.188 (0.22)	-0.462 (0.30)	0.020 (0.01)	-0.1E-3 (0.9E-3)	0.112 (0.12)
	S-INEFF	0.966 (0.37)	0.411 (0.51)	-0.040 (0.02)*	0.004 (0.002)*	0.016 (0.20)

Notes: Standard errors in the parenthesis/* denotes significance at 5%.

Concentration has a positive coefficient in thirty-eight out of forty five cases, where fifteen of these coefficients are statistically significant, while the coefficient of market share is in ten cases positive but only once statistically significant. This is consistent with the joint hypothesis that concentration proxies market power and that firms with more market power are less persistent in controlling cost. Our results are consistent with those of Berger and Hannan (1993, 1998), and Berger (1995a, 1997) that find support for Hicks 'quiet life' hypothesis.

As mentioned in the previous section, firms with market power purposely engage in some activities that raise cost more than revenues, and such activities will result in lower measured cost efficiency. Any gains from the positive profit structure relationship will tend to be offset by this cost increase. This may provide an explanation for the weak profit-structure relationship in the current test and in the previous literature as well.

5.7 Conclusions.

This study uses directly calculated measures of bank efficiency to distinguish among alternative explanations of the relationship between market structure and bank performance in the UK banking industry. An innovation of the current study is the utilisation of different efficiency measures to test the sensitivity of the results.

On average the findings appear consistent for the first decade: for most of the years there is a support of the structure performance hypothesis. The positive relationship between concentration and inefficiency measures supports the 'quiet life' hypothesis, offered in addition to the market power hypothesis. On the other hand, the relative market power hypothesis, which asserts that firms with larger market shares are able to exercise market power in pricing differentiated products, is not supported by the data.

Concerning the efficient structure explanation for the existence of the positive structure-profit relationship, the results are somewhat mixed. Examining the relation of inefficiency variables with profit, a statistically significant negative relationship between them is found in some cases. This constitutes an acceptance of the ES hypotheses. Nonetheless, when the necessary conditions for the ES hypotheses are tested, these results are not supported. In addition to the above, consistent with most of the literature [see Molyneux (1996)], the findings are not very robust and they are sensitive to the measure of performance used, and in some cases sensitive to the measure of X-efficiency.

Last, there are doubts on the ability to improve profits through efficiency or market power. Results from estimating equation (5.8) produced results with significantly low R^2s. In some cases the introduction of the efficiency variable raise this value substantially (in some cases by 0.20) indicating that to some extent efficiency does affect profits. However, the R^2s for most of the regressions are still below 0.30 and sometimes very close to zero. Apparently this suggests that the variation in profits is due to other factors as well, such as portfolio choices and regulatory changes.

To conclude, there is evidence that UK data on average support the structure conduct performance hypothesis and its 'quiet life' addendum, while there is little support for the relative market power hypothesis. In addition, as far as the efficient structure hypotheses are concerned, the data support only some of its conditions. Comparing our results to other European studies, we find that they are in line with those of Vennet (1993) and Molyneux and Forbes (1995) who also find a support for the SCP hypothesis, but in contrast with results of Goldberg and Rai (1996) whose data support the RMP hypothesis.

CHAPTER 6: The Effects of Regulation on the Performance of UK Banking Institutions

6.1 Introduction

In this chapter our goal is to determine whether deregulation altered the profits, costs, and efficiency of the UK banking institutions, for the period 1983 to 1995. The motivation for this study comes from the notion that the deregulation of the banking industry that took place during the 1980s all over the world had an effect on the performance of banking institutions. Nevertheless, these effects have only been examined empirical by a handful of studies, none of which focuses on to the UK banking market, to the best of our knowledge.

The deregulation process emerged primarily by the globalisation of financial services; policy makers recognised that existing regulation was largely set up for the needs of the past, and hence it could no longer satisfy present financial needs. Deregulation has changed the nature of banking and, consequently, its production and cost structure. Banks play a vital role in the economy; their efficiency and productivity have significant consequences across a wide range of non-financial firms and industries. New regulatory requirements try to capture all risks associated with banking markets, such as credit risk, market risk as well as the risk of default. At the same time they try to promote competition in order to make financial institutions more efficient and productive.

Section 2 reviews the major regulatory tools utilised in the banking industry together with the associated literature. Section 3 discusses the motivation for the deregulation process that took place during the last two decades. Its also provides a review of the studies that have attempted to quantify these effects on banks' performance. Section 4 outlines the methodology applied, whereas the next section summarises the data utilised. Section 6 discusses the empirical results, and section 7 concludes.

6.2 Regulatory tools

There are two types of regulation traditionally proposed in recent regulatory theory: regulation that affects directly the structure of the industry and regulation that affects the behaviour of the industry participants. These are the structural and conduct regulation, respectively.

Structural regulation aims at regulating the actual structure of the industry. It includes functional separation of institutions, i.e. separation between commercial and investment banking. In many countries, there is a separation between banks of different types according to their functions. Consequently, a commercial bank needs to be separated from an investment bank. Also, there are entry requirements and discriminatory rules concerning foreign banks and investors. With structural regulation there are certain entry requirements which do not allow free entry and exit from the system. This also affects the presence of foreign firms, the size of banks, the frequency and type of mergers and acquisitions and the nature of the products that will be offered.

Conduct regulation takes the form of direct restrictions on assets and liabilities, rules relating to information disclosures, credit ceilings, limitations on branching and the determination of fees, commissions and rates on assets and liabilities. These kinds of regulation are expected to provide banks with the incentive to overemphasise competitive tools, which are not restricted [Gual and Neven (1993)].

The major regulatory tools in use in the banking industry could be classified into five broad types [Freixas and Rochet (1998)]:

1. Deposit interest rate ceilings.
2. Entry, branching, network, and merger restrictions.
3. Portfolio restrictions.
4. Deposit insurance.
5. Capital requirements.

Below we give a brief description of each type of regulatory tools.

6.2.1 Deposit rate ceilings

Deposit rate restriction was implemented in many countries during the 1970s and early 1980s. In the US it is known as *Regulation Q* and was imposed until 1986. In Europe, it still exists in several countries, such as Germany and France, and has been lifted only recently by others.

Deposit rate ceilings emerge from the belief that unrestricted competition among banks would increase deposit rates, and in turn bank costs. To compensate for the increase in their cost, banks would shift their portfolios to higher yielding assets. purchased at the expense of higher risk. Accordingly, deposit rate ceilings were originally imposed to limit inter-institutional competition for deposit as a means for reducing deposit costs, and therefore contributing to the financial integrity of depository institutions. An opposing view suggests that ceilings might increase bank risk by encouraging non-price competition for deposits during periods of fluctuating interest rates. This reduces capability of banks and increased exposure to disintermediation in periods of rising market interest rates [Benston (1964), Cox (1966), Klein (1971), Startz (1979), and Mingo (1980)].

The empirical literature on the effect of deposit rate restrictions is quiet extensive. A portion of studies examines the effect of these restrictions on banks' risk. In summary, early studies on the deposit rate-solvency issue suggest that bank risk remains unaffected [Benston (1964), Cox (1966), Koehn and Stangle (1980), Smirlock (1984),] or increases by deposit rate ceilings [Startz (1979), Mingo (1978, 1980)]. In either case the implication is that the relaxation or elimination of deposit rate ceilings does not impair the solvency of the banking system. Later studies examine the impact of abolition of restrictions. The results here are somewhat mixed. Ahorny *et al.* (1988) claim that there is a decline in systematic risk, whereas non-systematic and total bank risk rise. On the other hand, Allen and Wilhelm (1988) find no evidence of an increase in bank risk. Finally, results by Bundt *et al.* (1992) indicate an increase in both systematic and non-systematic risk.

Some studies attempt to determine to what extent, if at all, deposit rate restrictions impact adversely on commercial banks, and whether they constitute a significant cause for bank failures. Cebula and Saltz (1994) and Ostrosky (1997) suggest that Regulation Q, among other factors[73], has been a statistically significant contributor to the growth of commercial banks failure rate in the US during the early 1980s.

Another issue covers by the literature covers is the wealth effect of deposit rate regulation. Ceilings act as a government fixing agreement, which in turn acts as a tax on depositors; the existence of deposit restrictions acts as a transfer of wealth from depositors to the owner of the institution[74] and, thus ceilings impose losses to depositors [Pyle (1974), James (1983)]. On the other hand, decreasing costs of resources for banks may entail a decrease in the rate they charge to borrowers, as well as an increase in the quality of financial services [Freixas and Rochet (1998)].

James (1983), by use of event study methodology, finds that significant intra-industry wealth transfers result from the imposition of deposit rates ceilings. Wholesale and retail commercial banks have experience an increase in their value with the removal of ceilings on the CDs, suggesting that ceilings on these accounts have acted as a tax on earnings. On the other hand, on the announcement of ceiling changes on consumer accounts, retail banks experience a significant decrease in their value, consistent with the hypothesis that ceilings applied to consumer accounts provide these institutions with a subsidy.[75] Allen and Wilhelm (1988) find that following the abolition of deposit rate restrictions the competitive structure of the depository institution changed, with the Federal Reserve System banks profiting at the expense of other parties in the industry (non-member banks and S&Ls). In an examination of the impact of deposit rate ceilings in the French banking industry, Dermine and Hillion (1992) find evidence that demand deposits provide rents when zero interest regulation is applied.

[73] Other statistically significant factors that are found to affect bank failure in the US are the recession during the early 1980s, the real cost of funds to banks, as well as the Tax Reform Act of 1986.
[74] Whether ceilings achieve a wealth transfer from the depositor to the owner of the institution depends on the elasticity of deposit supply, and the extent to which ceilings can be circumvented through the use of non-interest payments.
[75] Peltzman (1976) suggests that if producers differ in their cost structures or if differences in product demand exist then certain firms in the regulated industry may earn subsidies while others are taxed by regulation.

6.2.2 Entry, branching, network, and merger restrictions

To discourage concentration, US authorities have always imposed restrictions on branch and network banking[76] until mid 1990s, unlike most banking in the industrialised world[77]. This was seen as one mean for offsetting oligopolistic tendencies arising from a regulated sector[78]. Nevertheless, the recent literature suggests that these restrictions have increased bank instability by limiting the ability of banks to diversify their loan portfolios and deposit liabilities [Calomiris (1992, 1993)]. They have also reduced the efficiency of average bank assets and increased intermediation costs, primarily by limiting better-managed banks to gain market share [Jayarante and Strahan (1999)]. Moreover, geographic restrictions probably increased market power by limiting entry [Evanoff and Fortier (1998)].

During the 1980s and 1990s restrictions on banks' ability to expand geographically have been relaxed with a series of removals of restrictions on interstate and intrastate banking. With *Riegle-Neal Interstate Banking and Branching Efficiency Act* (IBBEA) of 1994 interstate banking has been permitted in almost all States. Studies on the US market suggest that due to deregulation, there was an increase in Merger and Acquisition (M&A) activities in States after they joined interstate banking agreements [Jayarante and Strahan (1996)]. Further, with the relaxation of state-wide branching restrictions, there has been an increase in the number of new branches as well as in the number of entry into local markets via *de novo* banks [Amel and Liang (1992), Calem (1994)]. In addition, following the IBBEA adoption, studies find large positive bank stock returns, suggesting that take-over deregulation enhances equity values [Brook *et al.* (1998)].

Europe has also been undergoing deregulation. The major step towards the elimination of entry barriers in the banking sector has been the implementation of the

[76] Geographic restrictions that prevent banks from branching outside a single county act as a form of entry restriction.
[77] The *McFadden Act* (1927) permitted branching for national banks but required them to obey state branching regulation. The *Douglas Amendment of the Bank Holding Company Act* (1956) gave each state the ultimate authority to approve entry by out-of-state banks. *Bank Merger Act* (1960) gave guidelines for mergers. The *Riegle-Neal Interstate Banking and Efficiency Act* (1994), changed the Bank Holding Company Act allowing adequately capitalised and well managed BHCs to acquire banks in other states.
[78] See Hefferman 1996, p.p.240-242.

Second Banking Directive (1993)[79]. The most important aspect of this Directive is the provision of a single banking license, which enables credit institutions incorporated in a Member State to be equally recognised throughout the EU by virtue of their home country authorisation. Further to the above, since early 1980s EU member countries have continued to deregulate their financial markets and foreign exchange controls, and this has give the opportunity to many foreign banks to operate in their markets. A study by Salomon Brothers (1990) suggests that the most noticeable characteristic of cross-border activity in European banking since early 1980s is large banks acquiring much smaller banks, although there have also been various large domestic bank mergers. Consequently, it can be argued that deregulation of the internal market has resulted in greater concentration and less competition in domestic markets since domestic banking firms have merged to improve their competitive position relative to potential EU foreign competitors.

Empirical literature suggests that geographic restrictions may have allowed some inefficient banks to survive. The abolition of these restrictions has allowed some previously prohibited M&As to occur, which may have forced inefficient banks to become more efficient by acquiring other institutions, by being acquired, or by improving management practices internally [Berger *et al.* (1999)]. In other terms M&As may allow the institutions to achieve a scale, scope, or mix of output that is more profitable, to improve X-efficiency through changing organisational focus or managerial behaviour, and as to improve the risk-expected return trade-off.

In terms of scale almost all studies that have utilised 1980s data suggest that there is no significant efficiency to be gained, while there are possibly slight efficiency losses to be suffered from M&A involving large banks [Berger et al (1997)]. The same results hold for scope and product mix efficiencies [Berger et al (1987), Hunter et al (1990), Noulas *et al.* (1993), Bauer *et al.* (1993), Clark (1996)]. Nevertheless, a recent study by Berger and Mester (1997) that has employed 1990s data finds substantial cost scale economies, of about 20% of costs for bank sizes ranging from $10 billion to $25 billion in assets.

[79] For details see Chapter 1, section 3.

The findings of X-efficiency studies indicate no or very little cost improvement from M&As during the 1980s [Berger and Humphrey (1992), Rhoades (1993), DeYoung (1997)]. Studies with early 1990s data generate mixed results. Rhoades (1998) finds modest efficiency gains, while Berger (1998) finds very little improvement in average cost efficiency for M&As of either large or small banks. A study on the European banking markets by Molyneux *et al.* (1996) finds evidence that M&As between large banks within domestic banking markets or across national borders can create substantial cost savings (or cost decrease) depending on the merger partners chosen.

6.2.3 Portfolio restrictions

Another form of regulation is portfolio restrictions imposed to banks as a mean of reducing banks' risk. These restrictions can take several forms: the most widely known is reserve requirements. The introduction of reserve requirements diminishes the business risk of the bank's assets and, as a result, lowers the risk of default [Crouhy and Galai (1991)]. Reserve requirements force banks to hold a fixed percentage of their deposits in the form of reserves earning no interest. Insofar as these requirements are binding, they act as a tax on deposits [Black (1970, 1975)]. Thus, adjustments in reserve requirements might be expected to influence the value of future implicit taxes on e institutions.

A question that remains open, though, is whether banks or customers pay the additional tax burden. Fama (1985) suggests that banks have a comparative advantage in monitoring at a lower cost than individuals; hence banks remain the cheapest source of funds for many borrowers even after shifting the deposit tax on them.[80] To test this hypothesis the author compares the average rates of interest on commercial paper and large CDs and finds no significant differences, suggesting that holders of large CDs do not bear the tax. This indicates that bank loans possess unique informational properties, and that the tax falls wholly on borrowers. Consistent with Fama's hypothesis are the findings of James (1987) and Cosimo and McDonald (1998).

[80] See Leland and Pyle (1977), Diamond (1984, 1989, and 1996), Chant (1987), Battacharya and Thakor (1993) for a detailed discussion on the importance of informational assymetries in financial intermediation.

If the Fama hypothesis is correct, then the tax on deposits falls wholly on borrowers (or on borrowers and depositors collectively), and shareholders do not bear any tax. Hence, a change in the tax rate should not affect bank share prices. Nevertheless, empirical literature that examines the impact of specific changes in reserve requirements, finds evidence that bank share prices do respond to such changes [Kolari *et al.* (1988), Slovin *et al.* (1990), Osborne and Zaher (1992), Kenneth and Madura (1996)]. These findings state that the tax falls at least partly on bank shareholders. The justification provided for these findings is that since reserve requirements force banks to maintain non-interest bearing funds, they essentially consist a tax which equals the opportunity cost of these funds. Since bank opportunity costs are affected by adjustments in reserve requirement, their perceived value should be affected as well.

Another form of portfolio restriction is the Glass-Steagall type of regulation. Under the requirements of the *Glass-Steagall Act* of 1933 US, commercial banks, are not allowed to hold corporate equities, leaving the underwriting of securities to investment banks.[81] On the other hand, in Europe commercial banks are universal and hence are allowed to hold demand deposits while dealing with corporate equities. Nowadays, the deregulation process has lifted many earlier restrictions, whereas in the US there is a controversy as to whether or not the Glass-Steagall Act should be abolished.

The reasoning behind the Act is that equity holding by banks may increase their risk exposure, and there might exist a conflict on interests. The latter may arise when a bank combines lending and deposit-taking with underwriting. The key difference between commercial and investment banks arises from loan-making activities of the former. If a firm has an adverse shock and the public remains unaware, a commercial

[81] The Glass-Steagall Act prohibits national banks from engaging in securities activities either directly or through affiliates. The Act was passed following the allegations of abuse by commercial banks, recorded in the Pecora Committee hearings, which conducted an investigation of misdoings of banks. Specifically, the First National Bank and The Chase Bank, in regard to their underwriting activity and their investment banking affiliates, were accused, among other things, of misrepresenting the quality of new securities they had underwritten; of using their commercial and investment banking operations to repackage bad commercial loans into securities and passing them into unsuspecting public investors; and of pressuring correspondent banks to accept these poor quality securities, thus reducing the smaller lender's ability to provide credit at a time when it was crucially needed.

bank may have an incentive to underwrite public issues on behalf of the firm and use the proceeds for repay earlier bank loans made to the firm. Is has also been argued that a commercial bank, unlike an investment bank, may have access to a large number of unsophisticated depositors [Krosnzer and Rajan (1996)]. Thus, conflict of interest may induce commercial banks to fool the public into investing into securities that will turn out to be of low quality.

Nevertheless, empirical evidence suggests that this is not the case. Several studies compare the performance of securities underwritten by commercial and investment banks in the US prior to the enforcement of the Act, and find no such evidence. Instead, investors appear to have rationally accounted for the possibility of conflict of interest and this appears to have constrained banks underwriting high-quality securities [Ang and Richardson (1994), Krosnzer and Rajan (1996, 1997), Puri (1994)]. Moreover, there is evidence that investors are willing to pay higher prices for securities underwritten by commercial banks than by investment banks [Puri (1996)]. Hence, empirical evidences regarding the conflicts of interest support the Glass-Steagall deregulation.

6.2.4 Deposit insurance

Banks are subject to runs because of the transformation services that they offer [Diamond and Dybving (1983), Freeman (1988)]: since banks transform short-term deposits into long-term loans, they are exposed to a systematic risk if depositors choose to withdraw their funds.[82] If the net realisable value of the bank's assets (loans) falls below that of liabilities (deposits) then the bank will be unable to fully service withdrawals. This generates a negative externality for the bank experiencing liquidity shortage, since it implies an increase in the bank's probability of failure. Moreover, this can generate a negative externality for the banking system in total, since investors may be forced to withdraw their deposits from financial trustworthy banks in anticipation of failure due to general inability to satisfy withdrawals. In such a case a bank run may develop into a bank panic, which is costly to the economy

[82] When depositors observe large withdrawals from the bank, they may fear bankruptcy and respond by withdrawing their own deposits.

because the interruption of bank services severely affects productive investment if assets are prematurely liquidated.

To avoid bank panics and the associated social costs, governments have established deposit insurance schemes. Under such schemes banks pay a specific premium to the deposit insurance company, and in exchange depositors have their deposit insured -up to a fixed limit- in case of bank failure. Deposit insurance mechanisms were first developed in the US, following losses arising from the massive bank failures in the early 1930s. Later they have been adopted by most developed countries, but under different schemes. In the US, all member banks of the Federal Reserve System (FRS) must join the Federal Deposit Insurance Corporation (FDIC) and pay an insurance premium. In turn, the FDIC uses the premia to purchase securities and provides banks with a stream of revenue. Through these funds it insures deposits up to $100,000. During the 1980s more than 2,000 depository institutions failed. As a result, in 1991 the Congress enacted a fundamental insurance reform for banks and thrifts, the *FDCI Improvement Act.* The Act requires the FDIC to use a 'least cost' approach for resolving bank failures. Riskier banks are required to pay higher insurance premiums to the FDIC.[83]

In the UK in the event of failure depositors will have 90% of deposits refunded, up to a maximum payout of £18,000. The scheme is financed by a flat-rate contribution by banks in proportion to their deposits[84]. For most countries in the EU compensation per person of up to EURO 20,000 (approximately $21,000) is obligatory, although member countries are free to exceed this level if they wish. Many EU countries offer deposit insurance above that level; for example France has a maximum compensation per depositor of $81,650, Italy of $50,500, and Portugal of $45,175 .[85]

Credible deposit insurance minimises, one of the primary causes of a bank run since deposit convertibility is guaranteed and there is no need for withdrawals. So runs will not occur, and premature liquidation of assets is prevented ensuring that losses in the banking system due to panics will be avoided. Moreover, no sound bank would ever

[83]Hefferman, S. (1996) p.p. 235-238.
[84] Chapter 18 of M. Buckle and J. Thompson (1998)
[85] MacDonald, R. (1996), p.p. 17, and 22-24.

fail [Merrick and Saunders (1985)]. Further, Diamond and Dybving (1986) suggest that deposit insurance is the only known effective measure to prevent runs without preventing banks from creating liquidity; consequently, bank policies should be considered in the context of deposit insurance.

With the existence of deposit insurance banks have no incentives to take too much risk, and bank policy should be oriented towards minimising those incentives. Nevertheless, the provision of deposit insurance introduces moral hazard by freeing economic agents from the consequences of their actions [MacDonald (1996)]. Insured depositors no longer feel obliged to monitor the risk-taking behaviour of banks where their deposits are held. In such cases, depositors may choose depository institutions based on the interest rates they offer, rather than on their financial conditions. Further, incentives exist for banks to incur greater risk in an effort to earn higher expected profits, since they do no bear the cost in the case of an adverse event.

In a theoretical model, Freeman (1988) shows that moral hazard distorts the selection of bank's portfolio of liquid and illiquid assets even though all assets are free of default risk. This forces government to impose constraints that remove (or at least reduce) the effects of moral hazard from the selection of bank's assets and liabilities. Specifically, the model suggests that banking regulation such as required reserves of illiquid assets and limits or rates on returns can be justified as a response to moral hazard even in the absence of default risk.

Consistent with Freeman's notion that extra requirements are needed in the attempt to overcome moral hazard, some theoretical literature suggests that moral hazard incentives can be attenuated, and fail insurance pricing can be achieved most efficiently with the by capital requirements [Kupiec and O'Brien (1998), Gjerde and Semmen (1995)]. Others have argued that an increase in minimum capital requirements can create incentives for additional risk taking [Kim and Santomero (1988), Battacharya and Thakor (1993), Gennotte and Pyle (1991)]. Matutes and Vives (1995) indicate that with flat-premium deposit insurance maximum risk incentives exist even when there is no moral hazard, while Choury and Galai (1991) suggest that capital requirements and/or reserve requirements should complement fixed-premium deposit insurance.

Since the primary purpose of deposit insurance schemes is the prevention of bank runs, a part of the empirical literature examined the effect of deposit insurance on the failure of depository institutions. However, the evidence suggests the contrary. Cebula (1993) finds that federal deposit insurance has been a significant contributor to the S&L crisis in recent years. Wheelock and Wilson (1995) suggest that deposit insurance membership (banks members of the FDIC) increases the probability of failure. Saltz (1997) finds evidence of a long-run relationship between the bank failure rate and the deposit insurance provided by the goverment. Finally, Cebula and Belton (1997) indicate that the longer the extent of federal deposit insurance coverage, the higher the bank failure rate. The empirical results from the aforementioned studies provide support for the FDIC Improvement Act requirement of risk related deposit insurance premiums. Studies that examine the effect of the FDCI Improvement Act on the riskiness of the commercial banking system find a statistical positive relationship [Shiers (1994)].

6.2.5 Capital requirements

A bank's capital is required as cushion to cover up for losses that should be inherent by shareholders rather than depositors, as well as to finance the infrastructure of the firm [MacDonald (1996)].

The Bank of England first addressed the importance of capital adequacy. In 1974 a committee was set up and recommended the use of two ratios, the free resource (gearing) ratio and the risk asset ratio.[86] These ratios were further revised and constitute the basis on which the Bank of England appraises the capital adequacy of deposit taking institutions throughout most of the 1980s.[87] The gearing ratio is the ratio of current liabilities to capital resources, excluding that part of capital utilised to finance infrastructure and other non-banking activities. Because it does not take into account the riskness of the bank's assets, it is barely referred to nowadays for supervisory purposes. The risk asset ratio refers to the risk of losses generated by the use of the assets of the institution to the capital, which is available to finance such

[86] see *Bank of England Quarterly Bulletin*, September 1975.
[87] see "The measurement of capital" in the *Bank of England Quarterly Bulletin* of September 1980.

losses[88]. This ratio has become the key ratio in monitoring capital adequacy at international level.

In the US, the first capital requirements were introduced in 1981. Capital guidelines required banks to hold a flat minimum percentage capital against all assets. With the growth in off-balance sheet activities, there was a need for re-adjusting capital requirements to include new kinds of risks to which the banks were exposed. During 1986 the US Federal banking agencies proposed a risk-based capital measure that would take into account differences in risk among financial institutions, as well as off-balance sheet items.

In 1987, new risk-based capital rules were proposed based on a joint US/UK agreement in an attempt to encourage co-ordination among supervisory authorities from major industrial countries. Nevertheless, the scope of the international effort was further expanded with the Basle Committee on Banking Regulation and Supervisory Practices.[89] With the Basle agreement in July 1988, this Committee has proposed a new system of risk weightings and capital definitions, which requires from banks to achieve a certain ratio of capitalisation for risk-weighted assets. Specifically, all assets are divided into categories, and each category is given a weight according to its perceived related risk. All riskless assets, i.e. government bonds, are excluded from the risk asset base, while commercial loans are counted as its entirely as risky assets. Next, each asset is multiplied by its related risk factor, and then the total risk of the risk-adjusted assets is related to the bank's capital.[90] The capital itself is classified into two categories: *Tier I*, consists of equity capital and published accumulated profits; and *Tier II* consists of medium- and long- term subordinated debt instruments, general provisions and unpublished profits. The ratio of capital to risk-weighted asset must continuously be at least 8%. In the case where a bank does not satisfy the above requirements, the regulator must ask the bank to be recapitalised to the minimum level, and if this does not occur the regulator must either sell or

[88] The objective of these two capital ratios is first to ensure that the capital of an institution is regarded as acceptable by its depositors and other creditors (gearing ratio), and second to test the adequacy of capital in relation to the risk of losses which may be sustained (risk asset ratio).

[89] This Committee, set up, in 1975 is established by the central bank governors of Belgium, Canada, France, Germany, Italy, Japan, Luxembourg, Netherlands, Sweden, Switzerland, UK, and US.

[90] Off-balance sheet items are firstly transformed into balance items by the use of conversion ratios, and then the risk weights are applied to the converted values.

liquidate the bank, depending on which is cheaper. These requirements have come into effect in most major industrial countries by the beginning of 1993.

The Basle risk-asset ratio has been criticised on the grounds that is suffers from several drawbacks. First, there is almost no measure of portfolio risk. Second, it is based on historical cost accounting, so that assets are unlikely to reflect their true market value. Third, because it is based on balance sheet information it does not adjust quickly to new information. Finally, it is designed to measure banks' creditworthiness and says very little about the bank's potential exposure to short-term liquidity problems [Tirole (1993)].

In an attempt to capture market risk, the Basle Committee in 1996 has amended the Basle Accord by setting new minimum capital requirements for a bank's market risk exposure. These requirements have come into effect in 1998. Under these new requirements a bank must separate its long-term investments from its trading book, where trading book refers to the position taken with a view to re-sale or short-term profit[91]. In addition, in 1997 the Committee has published a set of principles for the assessment and management of interest rate risk.

Further to the above the Capital Adequacy Directive (CAD) has been implemented in EU member countries in the beginning of 1996. The Directive sets out the minimum capital requirements for credit institutions and investment firms for the market and other associated risk. The CAD has also been amended to come in line with the Basle amendment.[92]

Turning now to the literature on capital adequacy regulation, it is clear that the most important question is whether increased capital standards increase or decrease banks' risk. Koehn and Santomero (1980), Kim and Santomero (1988), and Rochet (1992) use a mean-variance framework and show that improper risk weights may increase the riskness of banks. They also suggest that the probability of bank's failure (insolvency

[91] The bank's trading book position in debt, equity and derivative securities as well as commodity and foreign exchange positions held by the bank are subject to market risk capital requirements.
[92] See chapter 2 section 3 for a detailed analysis.

risk) may increase as a result of an increase in the capital standards. Nevertheless, they show that this distortion disappears when the regulators use market-based measures of risk.

Following a different approach Furlong and Keely (1989) and Keely and Furlong (1990) examine the case of a value-maximising bank. They show that a bank with publicly traded stock that maximises its portfolio will never increase portfolio risk as a result of increased capital standards. The key feature of their analysis is that under flat-deposit insurance, the bank will take more into account its prospective portfolio losses after an increase in capital. These results contrast evidence from mean-variance utility maximisation models. Gennotte and Pyle (1991), again with the use of a value-maximising model, show that both the portfolio risk and the possibility of bank failure may increase after an increase in capital requirements.

Recently, Blum (1999) indicates that the previous literature concentrates on the static effects of capital adequacy, and does not take into account the dynamic perspective of banking business. He shows that in a dynamic model with incentives for assets substitution, capital requirements might actually increase risk because under binding capital regulation there is a reduction in bank's profits. If bank's future profits are lower then there is no incentive for banks to avoid default. Further, these requirements raise the value of the bank's equity (leverage effect). Hence the bank must invest in more profitable -but more risky- assets, and for raising the amount of equity tomorrow it might be optimal to increase risk today.

From the above, we can say that there is no consensus in the theoretical literature on the impact of increased capital requirements on portfolio and insolvency risk. The results of the empirical literature are contradicting. The empirical examination merely tests whether risk-based capital requirements induce banks to shift their portfolios away from highly to lower risk-weighted assets. Evidence from the US suggests that banks have dramatically adjusted their portfolios towards assets in the low weight class (government securities and home mortgages) and away from high weight class (business and commercial loans). Some studies suggest that this alteration came as a result of the changes in capital requirements [Breeden and Isaac (1992), Haubrich and Wachtel (1993), Thakor (1993, 1996)]. Nevertheless, the conclusions as to whether this shift from one class to the other has been caused by changes in the capital requirements are somewhat controversial. Relevant studies provide at least more two different explanations for the reallocation of bank's portfolio. The first one is the implementation of leverage requirements by US regulatory authorities that mandates banks to hold capital of at least a certain percentage of unweighted assets [Syron and Randall (1992), Furlong (1992), and Shrieves and Dahl (1993)]. Second banks voluntarily have reduced their risk by having safer portfolios in 1990s as a result of the loan losses and commercial estate problems during the 1980s [Hancock and Wilcox (1992), Berger (1995b)]. On the other hand, UK banks did not alter their portfolios. Instead, they have achieved adjustments in their capital ratios primarily by directly boosting their capital [see Ediz, Michael and Perraudin (1998)].

To conclude we can say that capital is widely believed to reduce the bank's incentive to choose riskier asset. Nevertheless, some suggest that capital requirements may be ineffective to act as a controlling tool of bank risk, and might even influence the bank to choose riskier assets, while others have argued that capital requirements are less effective in determining bank risk than it is widely believed.

6.3 Deregulation and Bank Performance

In the previous section we have examined different regulatory tools together with the relevant literature. Going through the regulatory requirements, there is evidence that in most countries the business of banks did not change significantly from the late 1940s to the 1970s. The major reason relates to the strict regulatory framework followed by each country. However, in the 1980s we have witnessed a revolution in both banking practice and banking theory, due to an international deregulation pattern that has changed the nature of banking, and consequently its production and cost structure. Deregulation aims at liberalising the provision of services and making easier the entry of new banks in the market to increase market competition.

This deregulation pattern have emerged mainly by two mutually reinforcing causes:

1. increasing internalisation of financial and non- financial sectors,

2. existing regulation were largely set up for needs of the past, hence they could no longer satisfy the present financial needs.

For a long time bankers have concentrated on the question of how to circumvent regulation and this search has led them outside national borders towards internalisation. Taking into consideration the particular features of banking, national regulation of banking industry is very difficult. Border controls and transportation cost do not constitute any barriers because unlike goods, money can be transported virtually without any cost, and to some extent, invisibly. When all countries were heavily regulated, this did not constitute a problem. However, as soon as some countries opened their financial markets international competition occurred, which eventually led to a competition among the regulators. Several regulators saw the dangers of business outside domestic financial centres, so the need for regulatory adaptations was deemed necessary. In addition, the internalisation of financial transactions made it inevitable to establish an extensive prudential practice to ensure that bank's foreign operations do not escape from supervision.

Simultaneously with the search for circumventing regulation during the 1980s, there has been an increase of globalisation of the world economy. The major developments that have contributed towards internalisation can be summarised as follows. First, the growth of international trade outpaced production growth, and over time national economies have become more trade dependent. Also, reduced capital controls in certain countries made it possible for investors to take advantage of international differences in performance, and freedom to enter and exit home and foreign currency markets. Furthermore, the benefit of diversification of foreign assets in reducing the overall portfolio risk is dictated by modern portfolio risk theory. Finally, in the last two decades increasing shares of private wealth are managed by institutional investors, who seek to diversify risk though international portfolio management. All these factors made reduced controls and regulation both necessary and acceptable.

Deregulation transformed the banking industry. Geographical barriers have been lifted, both in Europe and US, and competition and consolidation activities have risen considerably. There have been changes in capital requirements, interest rates ceilings, reserve requirements, along with changes in technology. These fundamental changes are likely have had important consequences on industry costs, efficiency and profits, and the economic environment. Nevertheless, quantification of these changes has been fairly elusive.

Berger and Humphrey (1992) find that during 1980-1984 banks experienced increased cost, primarily due to the deregulation of deposit rates that resulted in an increase of competition in the deposit market. The increase in cost was larger for smaller banks, which might indicate that smaller banks rely more heavily on deposits in contrast to the larger banks. On the other hand, during 1984-1988 the authors find a decrease in the average cost for all bank sizes, suggesting that the effect of deregulation was exhausted.

Berger and Mester (1997) compare cost, profit efficiency and productivity during 1980s and 1990s. Their results suggest that large banks have shown a considerable decline in profit efficiency due to changes in the bank business environment, as total cost has increased during both periods due to changes in the business environment. On the other hand, profits increase during both periods, with the entire change reflecting increased profit productivity. Contrary to the above results, Humphrey and Pulley (1997) find that large banks have experience a rise in their profits from the 1977-1981 to the 1981-1984 period, due to a shift in the profit function and to changes in the business environment. Small banks are found to increase slightly their profits from one period to the other. The increase in profits continues and in the post-deregulation period for both groups of banks, but can be essentially attributed to changes in the business environment.

Finally, in a recent study Berger and Mester (1999) use data from the 1990s and find evidence of a decline in cost productivity and any increase in profit productivity. The authors suggest that banks nowadays provide a variety of new financial services, such as mutual funds and derivatives. At the same time they provide higher quality services, with extensive ATM networks and branching, and have increased the availability of debit and credit cards. All these services trigger additional costs, but at the same time generate additional revenues.

6.4 Methodology

Our goal is to determine whether there have been changes in bank cost, efficiency and profits for the period from 1983 to 1995 after the deregulation in the UK. Changes in cost and efficiency are reflected in the actual industry operating plus interest cost, whereas regarding profits, changes are reflected by standard bank profitability measures, namely return on assets and return on equity. Following changes in regulation, bank profits, costs and efficiency are mainly influenced by two factors:

- *Internal factors:* bank initiated adjustments in output prices and use of input, to altered regulatory structure.

- *External factors:* contemporaneous changes in banks' business environment, reflected in variations in output quantities, input prices, and other included variables.

To capture these effects on profits and costs, an indirect profit function, is estimated together with a cost function, to capture the effects on cost and efficiency. Further, the internal and external factors are separated to identify the potential effects of regulatory changes on banks' profits and costs efficiency.

To separate these two types of effects and determine both the relative importance of bank-initiated adjustments as well as the length of the adjustment period we have to make an assumption about the competitive structure of the industry. The standard methodology assumes perfect competition, where output and input prices are taken as given and banks' responses are limited to changes in input and output quantities. However, the assumption of perfect competition may not be valid in actual banking industry, where banks are able to exercise local market power for certain deposits and loan services. Furthermore, banks have the ability to differentiate intertemporally output prices among customer groups, across geographic areas and over time. Hence, we shall examine the role of both output prices and quantities based on an alternative assumption of imperfect competition. Under this approach banks have some control over output price, whilst output and input quantities comprise the external environment.

Based on the Humphrey and Pulley (1997) methodology, an indirect profit function is estimated, to separate the internal (bank-initiated) adjustments from external changes in banks' business environment. This separation will enable us to identify the major impact of changes in regulation on bank profit and technology. In addition, the same methodological specifications are going to be applied in each period's cost functions. The effects of regulation on bank efficiency are going to be estimated by utilisation of cost frontiers. This approach uses a cost function to measure how close a bank's cost is to what a best-practice bank's cost would be for producing the same output quantities when facing the same market prices of inputs.

6.4.1 An Indirect Profit Function under Imperfect Competition

In its general form, the indirect profit function is derived under the assumption of price taking behaviour in both input and output markets, i.e. perfect competition. However, the structure of banking industry is not characterised by perfect competition. As already mentioned, banks are able to exploit local market power for certain deposit and loan services, and have the ability to alter output prices among customer groups, and over time.

In the case of deposits, local market power evolves mainly from the desire of consumers to turn deposits into cash at low cost and, most importantly, to use local banks to obtain local acceptability of cheques for everyday transactions. In the case of loans, local market power rests on the role of banks as delegated monitors of customers: banks minimise transaction costs by monitoring loans and signalling in the context of asymmetric information. The generation of private information about loan customers is too costly to be reproduced by other lenders.

Although the competitive market assumption cannot be accepted, we cannot either accept the opposite model, monopoly. Under monopolistic conditions, input prices are determined exogenously, whereas output prices and quantities and determined endogenously. Thus, the profit function is insufficient to represent bank profit decisions.

Therefore, a model that describes bank's profits best must lie somewhere between the two aforementioned models. Under such a model banks can exercise a form of market power in choosing output prices [Humphrey and Pulley (1997)].

The problem now is to model the banks' mix of price taking and price setting behaviour. We start with the assumption that banks consider output as being exogenously given, as they have greater flexibility with regard to output prices than with output levels. "*Deposit output expands through growth in the local market, through mergers and acquisitions, or through use of purchased funds. All three offer only limited flexibility to bank managers in attempt to maximise profits over the intermediate term. Balance sheet equalities provide similar restrictions on loan outputs.*"[93] Consequently, we assume that banks do not focus on adjusting output quantities, but on negotiating output prices, to maximise profits.

More specifically, we assume that banks maximise profit, π, for given quantities of output, y, and input prices w, by choosing output prices p, along with input quantities x. The associated indirect profit function is given by:

$$\pi = P'Q = \pi(y, w, z) \tag{6.1}$$

where:

$P = (p, w)$,

$Q = (y, -x)'$ = vector of netputs

z = a set of variable expected to influence profitability.

The major benefit of this profit function is that it gives an appropriate specification under market power, as it reflects the bank's assessments of its competitive position and its assessment of the willingness of the customers to pay the prices that the bank wishes to charge.

[93] Humphrey and Pulley (1997) p.p. 8.

Here, we adopt the composite specification of Pulley and Breunstein (1992), and Pulley and Humphrey (1993, 1997). The general composite profit function associates a quadratic structure for output with a log quadratic structure for input prices. Output quantity and input prices are linked through interaction terms, and so separability is not imposed. The general composite profit function is represented by:

$$\begin{aligned}\pi^{(\phi)} = \{(&\alpha_0 + \sum\alpha_n z_n + \sum\alpha_i y_i + \frac{1}{2}\sum\sum\alpha_{ij} y_i y_j + \sum\sum\beta_{ik} y_i w_k) \\ &* \exp(\sum\delta_k w_k + \sum\sum\delta_{kl} w_k w_l)\}^{(\phi)} + \varepsilon\end{aligned} \tag{6.2}$$

where:

π = the profit measure,

$\alpha, \beta, \gamma, \delta$ = independent parameters,

w_k, w_l, = price of input k and l,

$y_i\, y_j$ = outputs i and j,

z = a set of variable that is expected to influence profitability,

ε = the error term,

and (ϕ) refers to the Box-Cox transformation[94] and represents an application of the "transform-both-sides" (TBS) to increase flexibility of the model. The application of Box-Cox transformation to the dependent variable and to the entire right hand side of the profit function, excluding the error term, preserves the composite structure of the model.

[94] The *Box-Cox* transformation was first introduced by Box and Cox (1964). In some cases, researchers may have a large amount of information about variables to be included in a particular relationship, but face lack of information about the precise form of the relationship. The Box-Cox transformation is used as a device to let the data determine what functional form is more appropriate. By transforming some of or all the variables in a relationship, a new set of functions is generated, with one particular member of this set being defined by specifying a particular value(s) of the transformation parameter(s). The functional form determined by the data is the functional form defined by the estimated value(s) of the transformation parameter(s).

Equation (6.2) can be viewed as a flexible second-order approximation to a wide class of specifications. In this model there is no logarithmic transformation with respect to input prices; while the logarithmic transformation applied to input prices makes the composite function easily restricted to be linear homogeneous in input prices (as required theoretically of an indirect cost function), the alternative indirect profit function does not require homogeneity. Therefore, we report the results without the logarithmic transformation.

As far as the variables in *z* are concerned, they are expected to influence profitability, and their effects are specified to be linear and neutral with respect to output and other variables. Although this specification allows to test the relationship empirically, adding a full set of second order terms for the variables in *z* together with the interaction terms with outputs is likely to increase multicollinearity. So it has been necessary to delete the second-order input price coefficient in equation (6.2) due to lack of input variability [see Berger, Humphrey and Pulley (1997), and Humphrey and Pulley (1997)].

Hence, the specification for the indirect profit function becomes:

$$\pi^{(\phi)} = \{(\alpha_0 + \sum \alpha_m z_m + \sum \alpha_i y_i + \frac{1}{2} \sum \sum \alpha_{ij} y_i y_j).\exp(\sum \beta_k w_k)\}^{(\phi)} + \varepsilon \qquad (6.3)$$

By replacing output quantity with output prices in equation (6.3) the specification of the model will become a standard profit function. Nevertheless, since banks are viewed as price takers in output quantities and input prices, and price setters in output prices and input quantities, the introduction of output prices in the model should not contribute substantially to the model's explanatory power. The reason is that the effect of output prices is assumed to be reflected through estimated coefficients. And last, but not least, output prices are difficult to be measured.

6.4.2 Estimating the effects of changing regulation on bank's profits

The composite profit function (6.3) is separately estimated for three-time periods:

1. the pre-deregulation period, 1983-1987- *Period 1*
2. the concurrent period, 1988-1992 - *Period 2*
3. the post regulation period, 1993-1995 -*Period 3*

Following Berger and Mester (1997a), Humphrey and Pulley (1997), and Berger and Mester (1999) we decompose the profit index into technological changes, reflecting bank initiated adjustments, and changes in the business environment. Accordingly, the following ratio is estimated:

$$\pi_{t+1} / \pi_t = \beta_{t+1} X_{t+1} / \beta_t X_t = \frac{(\beta_{t+1} X_t)(\beta_{t+1} X_{t+1})}{(\beta_t X_{t+1})(\beta_{t+1} X_t)} = \frac{(\beta_{t+1} X_{t+1})(\beta_t X_{t+1})}{(\beta_t X_{t+1})(\beta_t X_t)} \tag{6.4}$$

where:

β_t = estimated parameters of the profit function at period t, where t = 1, 2, 3,

X_t = values of the variable at period t.

A change in β from period to period will reflect bank initiated adjustments in deposits and loan output prices and use of labour, and capital, in response to regulation. A change in X, value of the variables, from period to period represents changes in the business environment, reflected in variations in output quantities, input prices, and other included variable, thus structural adjustments to regulation.

Firstly, we estimate the ratio of predicted profits for banks (π_{t+1}/π_t) for periods 1 to 2 and 2 to 3. Normalising, period 1 equal to unity any divergence from this, comprise increase or decrease due to either changes in technology and/or changes in the business environment. However, the source of profit changes can be seen more clearly when the profit index is decomposed into the associated changes in technology and business environment. For example, from equation (6.4), the decomposition of the profit index for changes from period 1 to 2 is:

$$\pi_2 / \pi_1 = \beta_2 X_2 / \beta_1 X_1 \tag{6.5}$$

6.4.3 Specification of the cost function

The cost function relates variable costs to prices of variable inputs, the quantity of variable outputs and random error, as well as inefficiency. Mathematically, this can be represented as follows:

$$TC_{it} = C_t(Y_{it}, w_{it})\varepsilon_{it} \tag{6.6}$$

with its logarithmic transformation given by:

$$\ln TC_{it} = \ln C_t(Y_{it}, w_{it}) + \ln \varepsilon_{it} \tag{6.7}$$

where:

TC_{it} = total cost (operating and interest cost) for bank *i* at time *t*, and

$C_t(Y_{it}, w_{it})$ = a cost function at time *t* with output quantity vector *Y* for bank *i* at time *t* and input price vector *w* for bank *i* at time *t*, where $t = 1, 2, 3$.

The term ε_{it} is treated a composite error term, which accounts partly for inefficiencies and partly for random error.

A translog specification will be applied to estimate the cost function because as it is mentioned in the previous chapters, the translog specification overcomes the main drawbacks of the other major cost functions (see Chapter 3 and Chapter 4). Accordingly, when the minimum total cost function (cost frontier) is translog, then the bank's observed total cost is the following[95]:

[95] We estimate bank efficiency for our three sub-periods using the stochastic frontier approach (see chapter 4).

$$\ln TC = \alpha_o + \sum_{i=1}^{n} \alpha_i \ln w_i + \sum_{k=1}^{m} \beta_k Y_k + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij} \ln w_i \ln w_j +$$
$$\frac{1}{2}\sum_{k=1}^{m}\sum_{\zeta=1}^{m} \delta_{k\zeta} \ln Y_k \ln Y_\zeta + \sum_{i=1}^{n}\sum_{k=1}^{m} z_{ik} \ln w_i \ln Y_k + u + v \qquad (6.8)$$

where:

w_i, w_j, = price of input i and j,

Y_k Y_ζ = outputs k and ζ,

TC = total cost,

v_i = random error, where $v_i \sim N(0, \sigma_v^2)$,

u_i = inefficiencies, where $u_i \sim |N(0, \sigma_u^2)|$, and

α_i, β_k, γ_{ij}, δ_{kj}, z_{ik} are independent parameters.

6.4.4 Estimating the effects of changing regulation on bank's cost

The cost function (5.3) is separately estimated for three time periods, corresponding to pre-deregulation, concurrent, and post regulation. Following Berger and Mester (1997a) and Berger and Mester (1999), the total change in industry costs over time may be decomposed again into two elements. The first reflects changes in the productivity, which are changes in cost for given business conditions, and the second gives changes in cost due to business conditions themselves, which are simply the exogenous variables specified in the cost function.

Next, the following ratio is estimated:

$$C_{t+1} / C_t = \beta_{t+1} X_{t+1} / \beta_t X_t = \frac{(\beta_{t+1} X_t)}{(\beta_t X_t)} * \frac{(\beta_{t+1} X_{t+1})}{(\beta_{t+1} X_t)} \qquad (6.9)$$

where:

C_t = the estimated 'average-practice' cost function, which is estimated using all banks and embodies inefficiencies.

β_t = estimated parameters of the cost function at period t, where t = 1, 2, 3,

X_t = values of the variable at period t, including the inefficiency factor.

The first term on the right hand side of equation 6.10 represents productivity changes, i.e. changes in the productivity from period 1 to 2 is equal to:

$$\text{Productivity Changes} = \frac{(\beta_2 X_1)}{(\beta_1 X_1)} \tag{6.10}$$

The second term on the right-hand side of (6.9) represents changes in business conditions, i.e. changes in the business conditions from period 1 to 2 is equal to;

$$\text{Business Conditions Changes} = \frac{(\beta_2 X_2)}{(\beta_2 X_1)} \tag{6.11}$$

A change in β from period to period will reflect bank initiated adjustments in deposits and loan output prices and use of labour and capital, in response to regulation. A change in X (value of the variables) from period to period represents changes in the business environment, and thus structural adjustments to regulation. A change in the scale of operation will be included to a change in business conditions for the cost function, since output quantities are included in the exogenous variables of the cost function.

We estimate the ratio of predicted cost for banks (C_{t+1}/C_t) for periods 1 to 2 and 2 to 3. After normalising period 1 equal to unity, an increase or decrease is due to either changes in productivity and/or changes in the business environment. However, the source of these cost changes can be seen more clearly when the cost index is decomposed into the associated changes in technology and business environment. The results obtained from estimating equations (6.10)-(6.11) are presented in section 6.6.

6.5 Data Description

Following the major regulatory changes in the UK during the late 1980s and early 1990s (see Chapter 1) we consider the 1988 to 1992 period as the concurrent regulatory time interval. Accordingly, we have divided the data into three sub-periods:

1. 1983 to 1987 is considered as the *pre-deregulation period*,
2. 1988 to 1992 is considered as the *concurrent period*, and
3. 1993 to 1995 is considered as the *post-regulation period.*

The empirical analysis uses the same data set as the previous chapters. As far as the variables utilised is concerned, a brief description is given below.

For measuring the bank profits, two profitability ratios are used, the return on assets (ROA) and the return on equity (ROE). Return on assets is the ratio of net income divided by total assets and measures the efficiency with which total assets are employed within a firm. Return on equity is the ratio of net income to shareholders' equity, and it appraises the efficiency with which common shareholders' equity is being utilised within the firm. These are the most commonly used profitability measures in the previous literature.

The inputs selected are labour (x_1), and physical capital (x_2). Labour is measured by the total number of employees at the end of the time period, while physical capital by the net book value of total fixed assets. The price of labour (w_1) is derived by taking the total expenditure on employees and dividing it by the number of employees. A proxy for the price of capital (w_2) is calculated by dividing the total expenditure on fixed assets, with the net book value of fixed assets. The two outputs used in the current study, all measured in pounds, are Total Deposits (Y_1), and Total Loans (Y_2).

We also include a set of two control variables, contained in vector z, known within the banking industry to affect profitability. The first one is a risk measure (RISK), which is measured by the ratio of total liabilities to total assets. The second control variable is the size variable (TA) represented by the logarithm of total assets of each individual bank to control for differences in size and to capture the ability of banks to diversify.

6.6 Empirical Results.

6.6.1 Effects of deregulation on profits

To estimate the indirect profit function the non-linear least square estimation method is applied. The composite profit function (6.3) is separately estimated for three time periods noted above, pre-regulation, concurrent, and a post-regulation, allowing the coefficients to vary to reflect changes in profits, technology and business environment. Both ROA and ROE are utilised as depended variables, but because both estimates obtained are quiet similar, only those produced by ROE are presented.[96] Table 6-1 presents the coefficients obtained after the estimating the indirect profit function, equation (6.3), for each period.

Table 6-1: Parameter estimates for the three sub-periods.

Variables	**Post-regulation 1993-1995**	**Concurrent 1988-1992**	**Pre-regulation 1983-1987**
Constant	-0.366 (0.16)	-0.540 (0.71)	0.353 (0.19)
RISK	-0.172 (0.23)	-0.040 (0.05)	-0.381 (0.15)
LTA	0.045 (0.02)	0.039 (0.05)	0.9E-02 (0.01)
Y_1	-0.014 (0.02)	0.019 (0.02)	0.016 (0.03)
Y_2	0.014 (0.02)	-0.017 (0.02)	-0.015 (0.04)
Y_1Y_1	0.2E-02 (0.9E-02)	0.4E-02 (0.3E-02)	-0.6E-02 (0.1E-01)
Y_2Y_2	0.2E-02 (0.7E-02)	-0.2E-02 (0.4E-02)	-0.4E-02 (0.01)
Y_1Y_2	-0.5E-02 (0.02)	-0.3E-02 (0.6E-02)	0.8E-02 (0.02)
w_1	-0.7E-03 (0.6E-03)	0.9E-02 (0.3E-02)	-0.1E-02 (0.9E-03)
w_2	0.038 (0.07)	0.121 (0.35)	-0.479 (0.48)

Notes: (1) Standard errors of estimated parameters are shown in parentheses/ * significant with 95% confidence interval.

[96]A problem that we faced during the estimation procedure, was convergence of coefficients. Because of differences in the magnitude of the variables, deposits and loans were scaled down by a factor of -6.

From the estimated coefficient attained from equation (6.3) and the values of the variables we next estimate equation (6.4) to obtained changes in profit, technology and business environment. Table 6-2 presents the results.

Table 6-2: Changes in Profit, Technology, and Business Environment.

Changes in	**Period 1-2**		**Period 2-3**	
PROFITS	π_2 / π_1	0.980	π_3 / π_2	1.549
Technology	β_2 / β_1	0.841	β_3 / β_2	1.087
Business Environment	X_2 / X_1	1.166	X_3 / X_2	1.425

From Table 6-2 we observe that profits drop during the period when changes in the regulation took place, while there is a substantial increase afterwards. Specifically, the rate of actual profit of banks (π_{t+1}/ π_t) for periods 1 to 2 was 0.98 showing a two-percent decrease. Going to next period, the post deregulation period, the profits became 1.54, indicating a rise of 56 percent compared to the previous period. The finding are consistent with Berger and Mester (1997,1999) who find evidence that following the deregulation in 1980s US banks experienced improvements in profitability during the 1990s.

In order to see more clearly the underlying source of these profit changes, we must consider the decomposed profit index. The decomposition of the index shows changes in the technology, as reflected in the profit function coefficients β_t, and changes in the business environment observed in the profit function data X_t. Considering periods 1 to 2, if the business environment is held constant, then changes in technology would have generated a profit ratio of 0.84. However, actual profits indices were 0.98, implying that the business environment increases from 0.84 to 0.98, which is accomplished by a corresponding ratio of business environment index of 1.166. On the other hand, if technology is held constant the profit index would have rise by 0.16. Hence, the reduction in profit from period one to period 2 is the result of a change in technology ($\beta_2 \neq \beta_1$) by changing deposit and loan prices and the use of labour, and capital in response to changes in the regulation (and by changes in the coefficients of variables in z). The effect of a changing business environment in terms of deposit and loan quantities, input prices, and the variables in z, helps to reduce the decrease in profits from period 1 to 2. Nonetheless, the changes in technology have a greater effect on profit. It is clear that the new regulatory requirements, such as new capital and risk requirements, and the liberalisation of capital movements between countries has resulted in a decrease in industry's profits.

These results contrast the determinants of profits during period of regulatory changes with the post-regulation period. Here the profit index jumps by 54%, but this increase is entirely due to changes in business condition ($X_3 \neq X_2$). If technology is held constant, business environment would increase profit by 42%, whereas if business environment were held constant profit would increase only by 8%. The effect of technology was relative small from the second to the third period, and this suggests that benefits of banks' structural adjustments to deregulation have been essentially achieved by 1993. Only between periods 1 to 2 there was a strong positive effect on profits from changes in technology. However, we must note that during this period there was a recession in the UK economy (see Chapter 1, section 3). Hence the reduction in banks' profits may not be solely correlated with changes in the regulation.

The results obtained from the current study are consistent with those of Humphrey and Pulley (1997). Although their data sample is from a totally different market (US) they find that during the transition period, i.e. from the pre-deregulation to the concurrent period, the rise in profits has been the result of a shift in the profit function, of changing deposit and loan prices, and of changes in the use of labour, capital and funding inputs in response to deregulation. On the other hand, shifts in the profit function did not account for the rise in banks' profits for the post deregulation period. Instead, the authors find that this rise was due to higher levels of deposits, loans and other assets, and changes in input prices, i.e. technological changes.

6.6.2 Effects of deregulation on costs and efficiency

The cost function in equation (6.8) is estimated for the three-time periods, using the normal half-normal model. According to this model, random disturbances follow a normal distribution and inefficiencies, consistent with the stochastic frontier approach, follow a half normal distribution. A maximum likelihood procedure was used for the estimates. Table 6-3 shows the coefficients obtained after estimating the cost functions (6.8) for the three periods.

Table 6-3: Parameter estimates for the three sub-periods.

Variables	**Post-regulation 1993-1995**	**Concurrent 1988-1992**	**Pre-regulation 1983-1987**
Constant	2.702 (3.01)	2.569 (2.34)	3.423 (2.00)
W_1	-0.587 (0.99)	-0.267 (0.66)	-1.282 (0.75)
W_2	0.429 (0.59)	-0.787 (0.32)	0.162 (0.40)
Y_1	-2.654 (0.99)	-0.200 (0.33)	-0.361 (0.63)
Y_2	2.216 (1.10)	0.584 (0.41)	0.131 (0.70)
Y_3	(nil)	(nil)	(nil)
$w_1.w_2$	0.262 (0.23)	0.086 (0.10)	0.126 (0.13)
$Y_1.Y_2$	0.158 (0.40)	-0.376 (0.07)	-0.045 (0.12)
$Y_1.Y_3$	-0.009 (0.11)	-0.002 (0.01)	0.024 (0.27)
$Y_2.Y_3$	0.008 (0.12)	0.006 (0.01)	-0.008 (0.06)
$w_1.w_1$	-0.084 (0.23)	0.115 (0.12)	0.116 (0.19)
$w_2.w_2$	(nil)	-0.109 (0.04)	(nil)
$Y_1.Y_1$	-0.185 (0.26)	0.217 (0.05)	0.108 (0.09)
$Y_2.Y_2$	0.061 (0.15)	0.198 (0.03)	0.007 (0.05)
$Y_3.Y_3$	0.005 (0.004)	(nil)	(nil)

$w_1.Y_1$	2.363 (0.56)	0.236 (0.14)	-0.014 (0.22)
$w_1.Y_2$	-2.417 (0.58)	-0.276 (0.14)	0.140 (0.23)
$w_1.Y_3$	-0.049 (0.07)	-0.010 (0.16)	0.020 (0.02)
$w_2.Y_1$	-0.111 (0.15)	0.0446 (0.06)	-0.110 (0.10)
$w_2.Y_2$	0.131 (0.15)	0.0159 (0.06)	0.072 (0.10)
$w_2.Y_3$	-0.058 (0.04)	0.005 (0.01)	-0.020 (0.02)

Notes: (1) Standard errors of estimated parameters are shown in parentheses/ * significant with 95% confidence interval.
(2) Certain variables are excluded, due to high multicolinearity.

Table 6-4 presents the total changes in cost, productivity changes, changes in business conditions, and changes in average efficiency for pre-, concurrent, and post-deregulation periods from estimating equation (6.9).

Table 6-4: Changes in Cost, Productivity, Business Environment, and Efficiency.

Changes in	**Period 1-2**		**Period 2-3**	
Total Costs	C_2 / C_1	1.132519	C_3 / C_2	1.455199
Productivity	β_2 / β_1	0.848679	β_3 / β_2	1.038379
Business Conditions	X_2 / X_1	1.33445	X_3 / X_2	1.401414
Average efficiency	($x\text{-eff}_2$ - $x\text{-eff}_1$)	5%	($x\text{-eff}_3$ - $x\text{-eff}_2$)	-2%

As can be seen from Table 6-4 the total predicted cost of producing bank's output rose by 45% between the first period to the third. Cost rose by thirteen percent from period 1 to 2, and by another 32% from period 2 to 3. The increase in cost from period 1 to 2 justifies the reduction in profits for the same time interval. On the other hand, consistent with Berger and Mester (1997,1999), between period 2 to 3 there is an increase in both cost and profits, in approximately the same manner. This, in turn, suggests that profit margins of the industry have remained constant despite the increase in the competition from non-bank rivals and foreign competitors after the liberalisation of financial services. To gain a better idea on the driving forces of these increases in cost we decomposed the changes in cost to changes in productivity and changes in the business environment, and we consider changes in the average efficiency of the industry.

Considering first productivity, cost declines by an average of 15% from period one to 2, indicating that banks on the frontier have shifted on average towards less costly production, and at the same time the dispersion from this frontier has decreased. These results are in line with the changes in efficiency, from which for the same period we have an increase of 5%. It is clear in the light of the Single European Market and the new capital requirements, banks changed their use of labour and capital, as well as the prices of loans, deposits, and securities investment. Nevertheless, cost productivity decreased from period 2 to 3. This comes as a surprise since the effect of regulation would be expected to reduce cost in light of a more competitive environment. These findings are consistent with efficiency estimates, as well as with the profits results. As we saw from the previous section structural adjustments to deregulation were essentially achieved by 1993, as the effects of banks' initiated adjustments were relatively small for the last sub-period.

Looking at the business condition changes in Table 6-4, we can see that they generally worsened from early 1980's to mid 1990's. It is obvious that the increasing patter cost followed over the last thirteen years has been due to changes in the business environment.

Considering now efficiency, the first period average is 73%, for the next sub-period it amounts to 78%, and for the last period to 76%. From period 1 to 2, there is an increase in efficiency of five percent, suggesting better utilisation of inputs. However, after the deregulation there is a slight decrease of 2%. Nevertheless, the efficiency estimates give a measure of the dispersion of the banks from the best-practice frontiers for the three sub-periods, where these frontiers are supposed to be different for each period. The decrease in efficiency estimates for period three may not indicate that banks are using more resources to produce a given level of output than before. Instead, the decrease may be due to frontier improvement.

Possible explanations for this cost increase may be the competitive pressure which led banks to increasing output quality, which in turn raised costs. Furthermore, the increasingly competitive environment led to an increase in the number of mergers, and thus to a more concentrated market. This may have left banks temporarily in a disrupt production process and at higher cost levels. Also, due to concentration, as we have seen from the previous chapter firms with market power purposely may engage in some activities that increase cost more than revenues; such activities will result in lower measured cost efficiency, and in turn increase total cost. Another reason might be that new capital requirements change the structure of balance sheets resulting in an at least temporal cost increase. Apart from regulation, technology has launched new products and this has changed the banking business. However, their fixed costs may have not been yet recovered.

6.7 Conclusions

In this chapter we have investigated the changes in costs, profits and efficiency for the UK banking industry following changes in the regulatory framework. We based our analysis on the cost and profit concepts, which are superior to other measures (like ratios of cost to input or output to input), since they are derived from economic optimisation in response to relative prices rather than optimisation based on physical technology [Berger and Mester (1999)].

Following the relevant literature, we firstly focus on the major regulatory tools currently in use in the banking industry. Regulation could be classified into five broad types: deposit interest rate ceilings, entry, branching, network, and merger restrictions, portfolio restrictions, deposit insurance, and capital requirements. Deposit rate ceilings are imposed to limit inter-institutional competition for deposit as a mean of reducing deposit costs. On the other hand, branching and network restrictions together with merger restrictions are imposed in an attempt to limit concentration, leaving few banks with high market shares. Portfolio restrictions, deposit insurance and capital requirements are imposed as a mean of reducing bank's risk.

Secondly, we present a summary review of the studies on the effect of deregulation on bank's performance. The general conclusion from these studies is that both profit and cost increased with the new banking structure. An explanation for this phenomenon is the better quality along with more costly services they now provide.

As far as our results are concerned, we find that productivity and efficiency increases from pre-deregulation period to concurrent, whereas business conditions for banks' cost worsen from one period to the other followed by a reduction in banks' profits. Increasing competition, new competitors in the market, and new technology, have adversely affected costs and profits as well. In the 1990s structural adjustments to deregulation have been achieved, as the effects of banks' initiated adjustments were relatively small for the post-deregulation period. However, the business environment continues to affect bank cost. Surprisingly, although the cost continues to rise, there is an increase in profits as well. This, in turn, suggests that profit margins of the industry have increased despite the increase in the competition from non-bank rivals and foreign competitors after the liberalisation of financial services.

CONCLUSIONS

Motivated by changes in the structure of the banking industry that took place during the last two decades, the current study has attempted to assess their impact on the costs and profits of UK banking industry by use of a new data set covering the period from 1981 to 1995. Since the bulk of the literature concentrates mainly on US market, the results from this study contribute to our understanding of the effects of regulatory and market changes on the performance of banking institutions in a leading European banking industry. Moreover, by utilising alternative methodologies we provide an insight into the robustness of the findings to different methodological specifications.

In Chapter 1, we attempt to identify economic and market trends, as well as regulatory alterations that stimulated changes in the UK banking industry over the 1980s and 1990s. Chapter 2 provides the theory of modern banking firm and outlines the rational for banks and banking regulation. We also focus in the definitional and measurement problems of bank input and output, accompanied by an extensive discussion of scale and scope economies, X-efficiencies, and market structure relationship in banking, to give an insight to the empirical test that follows.

In Chapter 3 we estimate a cost function for UK banks using the translog specification and we investigate the presence of scale economies and cost complementarities. Motivated by the definitional problems on input and output specification in the current literature, two models are used: the first one classifies banks as intermediators of financial services, while the second is based on the value-added approach, where the identification of input and output is based on the share of value added. The results are found to be robust to the input-output specification. The findings indicate the existence of decreasing returns to scale, which implies the presence of diseconomies of scale for UK banks' during this period. Further, they suggest lack of cost complementarities at the beginning of the data period. On the other hand, following deregulation, there is evidence of scope economies in joint production in the beginning of the 1990s.

These results are somewhat different from previous findings in the literature, which, in turn, suggests that national markets and industries do not operate identically. Information obtained from estimating scale economies can be used to inform government policy by assessing the effect of deregulation on efficiency. Due to the existence of diseconomies of scale, mergers and acquisitions that lead to a more concentrated financial system may imply higher cost. Thus, government policies should be designed so as to control the growth of banks, as the latter may lead to operational inefficiencies. On the other hand, the existence of cost complementarities during the 1990s in joint production of loans and deposits, and loans and investments, shows that deregulation has benefited banks by increasing their penetration into new activities. This benefit is carried over to consumers in the form of lower prices, as competition increases between banks and other financial institutions.

In Chapter 4 we test the X-efficiency of UK banks by estimating an efficient frontier and by measuring the average differences between observed banks and banks at the frontier. Recent literature suggests that X-efficiencies account for 20% or more of cost in banking whereas scale and product diseconomies only count for less than 5% of cost [Berger *et al.* (1993)]. In the current study we employ the stochastic *econometric frontier approach* based on its advantages over the other frontier methods [e.g. Berger and Mester (1997a), Berger and DeYoung (1997), Bauer *et al.* (1998)]. To test the effect of this assumption on the robustness of the results we estimate three models with alternative distributional assumptions. Empirical findings are in line with those of the current literature and indicate that X-inefficiencies account for 20 to 30 % of costs in UK banking. Moreover, the findings suggest that choices made concerning measurement techniques make very little difference in terms of average industry efficiency or rankings of individual firms. Finally, in line with the results of Berg *et al.* (1992), Zaim (1995), and Rogers (1998b) our results suggest that in the beginning of the 1990s UK banks started to realise the benefits from deregulation, which has helped banks to reduce their inefficiencies level.

The question whether banks are efficient is of interest to managers and shareholders, as well as to customers and regulators. If these institutions become more efficient, then profitability will improve, greater amounts of intermediated funds will be available and better prices and service quality for consumers will be obtained. Thus, banks become more viable and, therefore, the possibility of failure or being subject to take-over decreases substantially.

In Chapter 5 we examine the relationship between market structure and profit relationship in the UK banking industry. Direct measures of both market structure and efficiency are included, while an innovation of the current study is the utilisation of different efficiency measures to test the sensitivity of the results. There is evidence that UK data support the structure conduct performance hypothesis and its 'quiet life' addendum, while there is little support for the relative market power and efficient structure hypotheses.

These findings reflect a setting of prices in more concentrated markets due to greater market power that is less favourable to consumers. They further indicate that banks with market power purposely engage in some activities that raise cost more than revenues, and such activities will result in lower measured cost efficiency. Any gains from the positive profit structure relationship will tend to be offset by this cost increase. This finding may explain the weak profit-structure relationship in the previous literature and in the current test. The policy implications arising from the results are that antitrust actions are likely to be socially beneficial, by moving prices towards competitive levels and allocating resources more efficiently.

In Chapter 6 we examine whether deregulation has altered the profits, costs, and efficiency of UK banking institutions. The motivation for this study comes from the notion that the global deregulation of the banking industry that took place during the 1980s is expected to have an effect on the performance of banking institutions. Nevertheless, these effects have been examined empirically by only a few studies, and no evidence is available for the UK. By dividing the data set into three sub-samples, namely pre-regulation, concurrent, and post-regulation periods, we find that productivity and efficiency increase from pre-deregulation period to concurrent, whereas business conditions for banks' cost worsen followed by a reduction in banks' profits. Increasing competition and new technology have adversely affected costs and profits. In the 1990s structural adjustments to deregulation has been essentially achieved, as the effects of banks' initiated adjustments are relatively small for the post-deregulation period. However, the business environment continues to affect bank cost. Despite the continuously increasing pattern of cost after deregulation, there is an increase in profits as well. This, in turn, suggests that profit margins of the industry have increased despite the rise in competition from non-bank rivals and foreign competitors after the liberalisation of financial services.

Finally, it is important to point out shortcomings in existing research that should be addressed and outline potential areas for future research. There is consensus in the literature that differences in frontier efficiencies among banks exceed inefficiencies attributable to incorrect scale or scope of output. Nonetheless, there is really no consensus on the preferred method for determining the best-practice frontier against which relative efficiencies should be measured. Given the existing parametric and non-parametric models the solution seems to be in the addition of flexibility to parametric approaches and the introduction of a random error into the non-parametric approaches. Furthermore, given the main methodological limitation of the existing cost and profit functions, more flexible functions should be derived in order to produce more accurate estimates. Considering the structure-profit relationship it is evident that market share, concentration and efficiency explain only partly the variation of profits. Hence, other factors such as portfolio choices and regulatory changes should be examined as well. Last, more research should be undertaken in the European banking industry, as most empirical work has until now concentrated in the US market.

BIBLIOGRAPHY

Abrams, B.A., and Settle, R.F., 1992, "What can regulators regulate? The case of bank entry", *Journal of Money, Credit and Banking*, 24(4), 511-518.

Aharony, J., Saunders, A., and Itzhak, S., 1985, "The effects of the International Banking Act on domestic bank profitability", *Journal of Money, Credit and Banking*, 17(4): 493-506.

Aharony, J., Saunders, A., and Swary, I., 1988, "The effects of DIDCMA on bank stockholders' return and risk", *Journal of Banking and Finance*, 12, 317-331.

Ainger, D., Lowell, K., and Schmidt, P., 1977, "Formulation and estimation of stochastic frontier production function models", *Journal of Econometrics*, 6, 21-37.

Akhavein, J.D., Berger, A.N., and Humphrey, D.B. 1997, "The effects of Megamergers on Efficiency and Prices: Evidence from a Bank Profit Function", *Journal of Industrial Organisation,* 12, 95-139.

Alhadeff, D.A., 1954, *Monopoly and Competition in Banking,* Berkeley: University of California Press.

Allen, L. and Rai, A., 1993, "Economies of scale and scope in international banking", *Journal of International Financial Markets,* 3(2): 1-31.

Allen, P.R., and Wilhelm, W.J., 1988, "The impact of the 1980 Depository Institutions Deregulation and Monetary Control Act on market value and risk: Evidence from the capital market", *Journal of Money, Credit and Banking*, 20(3): 364-380.

Allen, R.F., and Hagin, A.S., 1989, "Scale Related Efficiencies as a (Minor) Source of the Profits-Market Share Relationship", *Review of Economics and Statistics,* 71(3): 523-26.

Aly, H.Y., Grabowski, R., Rangan, N., and Rezvanian, R., 1990," Technical, scale, and allocative efficiencies in US banking an empirical investigation", *Review of Economics and Statistics*, 72, 211-218.

Amel, D., 1993, "State laws affecting the geographic expansion of commercial banks", Unpublished paper, Board of Governors of the Federal Reserve System.

Amoako-Adu, B., and Smith, B.F., 1995, "The wealth effects of deregulation of Canadian financial institutions", *Journal of Banking and Finance*, 19, 1211-1236.

Ang., J.S., and Richardson, T., 1994, "The underwriting experience of commercial bank affiliates prior to the Glass-Steagall Act: A re-examination of evidence for passage of the act", *Journal of Banking and Finance*, 18, 351-395.

Arrow, K.J., Chenery, Minhas, B., and Solow, R.M., 1961, "Capital-labour substitution and economic efficiency", *Review of Economics and Statistics*, 43, 225-250.

Athanassopoulos, A.D., 1995, "Multivariate and frontier analysis for assessing the market and cost efficiencies of large scale branch networks", Working Paper, University of Warwick, Coventry, U.K.

Athanassopoulos, A.D., 1997, Service quality and operating efficiency synergies for management control in the provision of financial services", *European Journal of Operational Research.*

Athanassopoulos, A.D., 1998, "Nonparametric frontier models for assessing the market and cost efficiency of large-scale bank branch", *Journal of Money, Credit and Banking,* 30(2): 172-193.

Avery, R.B.and Berger, A.N., 1991, "Risk-Based Capital and Deposit Insurance Reform", *Journal of Banking and Finance,* 15(4-5): 847-74.

Baldini, D., and Landi, A., 1990, "Economie di scala e complementarieta' di costo nell'industria bancaria italiana", *L' Industria*, No. 1, 25-45.

Baltensperger, E., 1972, "Economies of scale, firm size, and concentration in banking", *Journal of Money, Credit and Banking*, 4(3): 467-488.

Baltensperger, E., and Dermine, J., 1993, "European banking: Prudential and regulatory issues", in J. Dermine (eds), *European Banking in the 1990s*, Oxford: Blackwell Publishers, Second Edition, 17-38.

Bank of England, 1980, "The measurement of capital", *Bank of England Quarterly Bulletin*, September.

Bank of England, 1980, *Annual Report* for 1979/80.

Bank of England, 1981, "Foreign currency exposures", *Bank of England Quarterly Bulletin*, April.

Bank of England, 1981, *Annual Report* for 1980/81.

Bank of England, 1982, "The measurement of liquidity", *Bank of England Quarterly Bulletin*, July.

Bank of England, 1982, *Annual Report* for 1981/82.

Bank of England, 1983, "Competition, innovation and regulation in British banking", *Bank of England Quarterly Bulletin*, September.

Bank of England, 1983, *Annual Report* for 1982/83.

Bank of England, 1984, *Annual Report* for 1983/84.

Bank of England, 1985, "Off-balance sheet risks - Note issuance facilities/ revolving underwriting facilities", Banking Supervisory Division, April, BSD/1985/2.

Bank of England, 1985, *Annual Report* for 1984/85.

Bank of England, 1986, "Large exposures in relation to Mergers and Acquisitions", Banking Supervisory Division, February, BSD/1986/1.

Bank of England, 1986, "Subordinated loan capital issued by recognised banks and licensed deposit-takers", Banking Supervisory Division, March, BSD/1986/2.

Bank of England, 1986, "The measurement of capital", Banking Supervisory Division, June.

Bank of England, 1986, *Annual Report* for 1985/86.

Bank of England, 1987, *Annual Report* for 1986/87.

Bank of England, 1988, "Advertising for deposits", Banking Supervisory Division, April, BSD/88/1.

Bank of England, 1988, "Banking Act 1987, Section 16: Statement of principles", May.

Bank of England, 1988, "Underwriting exposures", Banking Supervisory Division, February.

Bank of England, 1988, *Banking Act Report* for 1987/88.

Bank of England, 1988," Implementation of the Basle Convergence Agreement in the UK", Banking Supervisory Division, October, BSD/1988/3.

Bank of England, 1989, "Loan transfer and securitisation", Banking Supervisory Division, February, BSD/89/1.

Bank of England, 1989, *Banking Act Report* for 1988/89.

Bank of England, 1990, "Implementation in the United Kingdom of the Directive on Own Funds of Credit Institutions", Banking Supervisory Division, December, BSD/90/2.

Bank of England, 1990, "Implementation in the United Kingdom of the Solvency Directive", Banking Supervisory Division, December, BSD/90/3.

Bank of England, 1990, *Banking Act Report* for 1989/90.

Bank of England, 1991, *Banking Act Report* for 1990/91.

Bank of England, 1992 & 1995, "Implementation in the United Kingdom of the Solvency Ratio Directive", Banking Supervisory Division, November 1992, BSD/1992/6, and March 1995 BSD/1995/1, (amendments to the 1990 paper).

Bank of England, 1992, "Implementation in the United Kingdom of the Directive of Own Funds of Credit Institutions", Banking Supervisory Division, January, BSD/1992/1, (amendment to the 1990 paper).

Bank of England, 1992, "Verification of interim profits in the context of the Own Funds Directive", Banking Supervisory Division, August, BSD/92/5.

Bank of England, 1992, *Banking Act Report* for 1991/92.

Bank of England, 1993, "Implementation in the United Kingdom of the Directive on the Consolidated Supervision of Credit Institutions", Banking Supervisory Division, December, BSD/93/1.

Bank of England, 1993, "Implementation in the United Kingdom of the Directive on the Monitoring and Control of Large Exposures of Credit Institutions", Banking Supervisory Division, October, BSD/93/2.

Bank of England, 1993, "Statement of Principles: Banking Act 1987, The Banking Co-ordination (Second Council Directive) Regulations 1992", May.

Bank of England, 1993, Banking Act Report for 1992/93.

Bank of England, 1994, "Subordinated loan capital issued by UK-incorporated authorised institutions", Banking Supervisory Division, May, BSD/94/3.

Bank of England, 1994, *Banking Act Report* for 1993/94.

Bank of England, 1995, "Implementation in the UK of the Capital Adequacy Directive", Supervision and Surveillance, S&S/1995/2, April.

Bank of England, 1995, "Implementation in the UK of the Capital Adequacy Directive", Supervision and Surveillance, December, S&S/1995/4 (amendments to S&S/1995/2).

Bank of England, 1995, "Implementation in the United Kingdom of the Capital Adequacy Directive ", Banking Supervisory Division, October, BSD/95/4.

Bank of England, 1995, *Banking Act Report* for 1994/95.

Bank of England, 1996, *Banking Act Report* for 1995/96.

Bank of England, 1997, *A stable financial system*.

Basle Committee, 1988, 'International Convergence of Capital Measurement and Capital Standards - The Basle Accord', July.

Bauer, P., 1990, "Recent developments in the econometric estimation of frontiers", *Journal of Econometrics*, 46, 39-56.

Bauer, P.W., and Ferrier, G.D., 1996, "Scale economies, cost efficiencies and technological changes in Federal Reserve payments processing", *Journal of Money, Credit and Banking*, 28(4): 1005-1037.

Bauer, P.W., Berger, A.N., Ferrier, G.D., and Humphrey, D.B., 1998, "Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods", *Journal of Economics and Business*, 50, 85-114.

Bauer, P.W., Berger, and Humphrey, D.B., 1993, "Efficiency and productivity growth in US banking" in M.O. Fried, C.A.K. Lovell and S.S. Schmidt (eds), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford: Oxford University Press.

Baumol, W.J., 1977, "On the proper test for natural monopoly in a multi-product industry", *American Economic Review*, 67, 809-822.

Baumol, W.J., Panzar, J.C., and Willing, R.D., 1988, *Contestable Markets and the Theory of Industry Services*, revised edition, New York: Harcourt Brace Jovanovich.

Bell, F.W., and Murphy, N.B., 1968, "Costs in commercial banking: A quantitative analysis of bank behaviour and its relation to bank regulation", Working Paper, Federal Reserve Bank of Boston.

Bell, F.W., and Murphy, N.B., 1968, "Economies of scale and division of labour in commercial banking", *Southern Economic Journal*, October, 131-139.

Bell, F.W., and Murphy, N.B., 1969, "Impact of market structure on the price of a commercial bank service", *Review of Economics and Statistics*, 51, 210-213.

Benston, G.J, Berger, A.N., Hanweck, G.A., and Humphrey, D.B., 1983, "Economy of scale and scope", *Proceedings of the Conference on the Bank Structure and Competition*, Chicago, Federal Reserve Bank of Chicago.

Benston, G.J, Hanweck, G.A., and Humphrey, D.B., 1982," Scale economies in banking: A restructuring and reassessment", *Journal of Money, Credit and Banking*, 16(1): 435-455.

Benston, G.J., 1964, "Interest payments on demand deposits and bank investment behaviour", *Journal of Political Economy*, 72, 431-439.

Benston, G.J., 1965a, "Economies of scale and marginal costs in banking operations", *National Banking Review,* 2(4): 507-549.

Benston, G.J., 1965b, "Branch banking and economies of scale", *Journal of Finance*, 20, 312-331

Berg, S.A., and Kim, M., 1994, "Oligopolistic interdependence and the structure of production in banking: An empirical evaluation", *Journal of Money, Credit and Banking*, 26(2): 309-321.

Berg, S.A., Forsund, F.R., and Jansen, E.S., 1991, "Technical efficiency of Norwegian banks: The non-parametric approach to efficiency measurement", *The Journal of Productivity Analysis,* 2, 127-142.

Berg, S.A., Forsund, F.R., Hjalmarsson, L., and Suominen, M., 1993, "Banking efficiency in the Nordic countries", *Journal of Banking and Finance*, 17, 371-388.

Berger, A. N., 1991, "The Profit-Concentration Relationship in Banking ", Working Paper, Board of Governors of the Federal Reserve System Finance and Economics Discussion Series: 176.

Berger, A.N, 1993, "Distribution free estimated of efficiency in the US banking industry", *Journal of Productivity Analysis*, 4, 261-292.

Berger, A.N., 1995a, "The profit relationship in banking - Tests of market-power and efficient-structure hypotheses", *Journal of Money, Credit and Banking*, 27(2): 404-431.

Berger, A.N., 1995b, "The relationship between capital and earnings in banking", *Journal of Money, Credit and Banking*, 27(2): 432-456.

Berger, A.N., 1998, "The efficiency effects of bank mergers and acquisition: A preliminary look at the 1990s data, in Y. Amihud and G. Miller (eds), *Bank Mergers & Acquisitions*, Boston: Kluer Academic, 79-111.

Berger, A.N., and Hannan, T.H., 1989, "The price concentration relationship in banking", *Review of Economics and Statistics*, 71, 291-299.

Berger, A.N., and Hannan, T.H., 1992, "The price concentration relationship in banking: A reply", *Review of Economics and Statistics*, 74, 376-379.

Berger, A.N., and Hannan, T.H., 1993, "Using efficiency measures to distinguish among alternative explanations of the structure-performance relationship in banking", Working Paper, Board of Governors of the Federal Reserve System, 1993-18.

Berger, A.N., and Hannan, T.H., 1997, "Using efficiency measures to distinguish among alternative explanations of the structure-performance relationship in banking", *Managerial Finance*, 23(1): 6-31.

Berger, A.N., and Hannan, T.H., 1998, "The efficiency cost of market power in banking industry: A test of the 'quiet life' and related hypotheses", *Review of Economics and Statistics*, 80 (3): 454-465.

Berger, A.N., and Humphrey, D.B., 1991, "The dominance of inefficiencies over scale and product mix economies in banking", *Journal of Monetary Economics*, 28, 117-148.

Berger, A.N., and Humphrey, D.B., 1992a, "Measurement and efficiency issues in commercial banking", in Z. Grilliches (ed.), *Output Measurement in the Service Sectors*, National Bureau of Economic Research, Chicago: University of Chicago Press, 245-279.

Berger, A.N., and Humphrey, D.B., 1992b, "Megamergers in banking and the use of cost efficiency as an antitrust defence", *Antitrust Bulletin*, 37(3): 541-600.

Berger, A.N., and Humphrey, D.B., 1997, "Efficiency of financial institutions: International survey and directions for future research", Working Paper, Board of Governors of the Federal Reserve System, 1977-11.

Berger, A.N., and Mester, L.J, 1997b, "Inside the black box: What explains differences in the efficiency of financial institutions?", Working Paper, Board of Governors of the Federal Reserve System, 1997-10.

Berger, A.N., and Mester, L.J., 1997a, "Efficiency and productivity change in the US commercial banking industry: A comparison of the 1980s and 1990s", Working Paper, The Wharton Financial Institutions Centre, University of Pennsylvania, 1997-5.

Berger, A.N., and Mester, L.J., 1999, "What explains the dramatic changes in cost and profit performance of the US banking industry?," Working Paper, Board of Governors of the Federal Reserve System.

Berger, A.N., and Udell, 1993, G.F, "Securitazation, risk, and the liquidity problem", in M. Klausner, and L.J. White (eds.), *Structural Change in Banking,* USA: New York University Salomon Centre, 227-291.

Berger, A.N., Bonime, S.D., Covitz, D.M., and Hancock, D., 1999, "Why are bank profits so persistent? The roles of product market competition, informational capacity, and regional macroeconomic shocks", Working Paper, Board of Governors of the Federal Reserve System.

Berger, A.N., Demsetz, R.S., and Strahan, P.E., 1999, "The consolidation of the financial services industry: Causes, consequences, and implications for the future", *Journal of Banking and Finance*, 23(2-4): 135-194.

Berger, A.N., Hanock, D., and Humphrey, D.B., 1993, "Bank efficiency derived from the profit function" *Journal of Banking and Finance*, 17, 317-347.

Berger, A.N., Hanweck, G.A., and Humphrey, D.B., 1987, "Competitive viability in banking: Scale, scope and product mix economies", *Journal of Monetary Economics*, 20, 501-520.

Berger, A.N., Humphrey, D.B., and Pulley, L.B., 1996, "Do consumers pay for one-stop banking? Evidence from an alternative revenue function", *Journal of Banking and Finance*, 20, 1601-1621.

Berger, A.N., Hunter, W.C., and Timme, S.G., 1993, "The efficiency of financial institutions: A review and preview of research past, present, and future", *Journal of Banking and Finance*, 17, 221-249.

Berger, A.N., Leusner, J.H., and Mingo, J.J., 1997, " The efficiency of bank branches", *Journal of Monetary Economics,* 40, 141-162.

Besanko, D., and Thakor, A.V., 1992, "Banking deregulation: Allocational consequences of relaxing entry barriers", *Journal of Banking and Finance*, 16, 90-932.

Bhattacharya, S., and Thakor, A.V., 1993, "Contemporary banking theory", *Journal of Financial Intermediation*, 3(1): 2-50.

Bhattacharya, S., Boot, A.W., and Thakor, A.V., 1998, "The economics of bank regulation", *Journal of Money, Credit and Banking*, 30(4): 745-766.

Black, F., 1970, "Banking and interest rates in a world without money", *Journal of Bank Research*, 1, 8-20.

Black, F., 1975, "Bank fund management in an efficient market", *Journal of Financial Economics*, 2, 323-339.

Blum, J., 1999, "Do capital adequacy requirements reduce risks in banking?", *Journal of Banking and Finance*, 23, 755-771.

Bond, R.E., 1971, "Deposit composition and commercial bank earnings", *Journal of Finance,* 26(1), 39-50.

Boot, A.W., and Greenbaum, S., 1993, "Bank Regulation, reputation and rents: Theory and policy implications", in C. Mayer and X. Vives (eds), *Capital Markets and Financial Intermediation*, Cambridge: Cambridge University Press, 262-277.

Bourke, P., 1989, "Concentration and other determinants of bank profitability in Europe, North America, and Australia", *Journal of Banking and Finance*, 13, 65-79.

Box,G. and Cox, D., 1964, "An analysis of transformation", *Journal of Royal Statistical Society*, Series B, 211-264.

Breeden, R.C., and Isaac, M., 1991, "Thank Basle for credit crunch", *Wall Street Journal*, 4.

British Bankers' Association, 1990, *Code of conduct for the advertising of savings and deposit accounts and money market accounts.*

Brook, Y., Hendershott, R., and Lee, D., 1998, "The gains from take-over deregulation: Evidence from the end of inter-state restrictions", *Journal of Finance*, 53(6): 2185-2200.

Brucker, E., 1970, "A microeconomic approach to banking competition", *Journal of Finance*, 25, 1133-1141.

Buckle, M., and Thompson, J., 1998, *The UK Financial System: Theory and Practice*, Manchester: Manchester University Press, 3rd edition.

Bundt, T., Cosimano, T.F., and Halloran, J.A., 1992, "DIDMCA and bank market risk: Theory and evidence", *Journal of Banking and Finance,* 16, 1179-1193.

Burgess, D.F., 1974, "Production theory and derived demand for imports", *Journal of International Economics*, 3, 105-121.

Cairncross, A., 1993, *The British Economy since 1945,* UK: Blackwell Publishers.

Calem, P.S. 1994, "The impact of geographic deregulation on small banks", *Business Review,* Federal Reserve Bank of Philadelphia, 17-32.

Calem, P.S., and Carlino, G.A., 1991, "The concentration / conduct relationship in bank deposit markets", *Review of Economics and Statistics*, 73, 268-275.

Calem, P.S., and Rob, R., 1996, "The impact of capital-based regulation on bank risk-taking: A dynamic model", Working Paper.

Calomiris, C., 1992, "Do 'vulnerable' economies need deposit insurance?", in P. Brock (ed.) *If Texas were Chile,* San Francisco: ICS Press, 237-350.

Calomiris, C., 1993, "Regulation, industrial structure, and instability in US banking: An historical perspective", in M. Klausner and L. White (eds.) *Structural Changes in Banking,* New York: New York University Press, 19-116.

Caves, R. and D.R. Barton, 1990, *Efficiency in the US Manufacturing Industry,* Cambridge, Mass: MIT Press.

Cebula, R.J., and Belton, W.J., 1997, "An empirical note on the impact of federal deposit insurance on bank failures in the US", *International Advances in Economic Research,* 3(3): 281-287.

Cebula, R.J., and Saltz, I.S., 1994, "On Regulation Q and the bank failure in the United States", *Rivisista Internazionale di Science Economiche e Commerciali,* 41(2): 185-191.

Chant, J., 1987, "Regulation of financial institutions - A functional analysis", *Bank of Canada,* Technical Reports No 45.

Christensen, L.R., Jorgenson, D.W., and Lau, L.J., 1973, "Transendental logarithmic production frontiers", *Review of Economics and Statistics*, 55, 28-45.

Christensen, L.R., Jorgenson, D.W., and Lau, L.J., 1975, "Transendental logarithmic utility functions", *American Economics Review,* 65, 367-381.

Clark, A.J., 1984, "Estimation of economies of scale in banking using a generalised functional form", *Journal of Money, Credit and Banking,* 16(1): 53-68.

Clark, A.J., 1986, "Single-equation, multiple-regression methodology: Is it an appropriate methodology for the estimation of the structure-performance relationship in banking?", *Journal of Monetary Economics*, 18, 259-312.

Clark, A.J., 1996, "Economic cost, scale efficiency, and competitive viability in banking", *Journal of Money, Credit and Banking,* 28(3): 342-366.

Conigliani, C., DeBonis, R., Motta, G., and Parigi, G., 1991, "Economie di scala e diversificazione nel sistema bancario", *Banca d' Italia*, Temi di discussione 150.

Cooper, J.C.B., 1980, "Economies of scale in the UK building society industry", *Investment Analysis*, 55, 31-36.

Cordell, L.R., and King, K.K., 1995, "A market evaluation of the risk-based capital standards for the US financial system", *Journal of Banking and Finance*, 19, 531-562.

Cosimo, T.F., and McDonald, B., 1998, "What's different among banks? ", *Journal of Monetary Economics*, 41, 50-57.

Cossutta, D., ML., D.B., Giannini, C., and Ugra, G., 1988, "Processo produttivo e struttura dei costi nell'industria bancaria italiana", in F. Cesarini, M. Grillo, M., Monti and M. Onado (eds), *Banca e Mercato a cura di*, Bolonga: Il Milano.

Cowell, R. J., and Davis, E.P., 1992, "Output, productivity and externalities - The case of banking", Bank of England, Working Paper Series No 3.

Cox, A., 1966, "Regulation of interest on bank deposits", Working Paper, Bureau of Business Research, University of Michigan.

Crouhy, M., and Galai, D., 1991, "A contingent claim analysis of a regulated depository institution", *Journal of Banking and Finance*, 15, 73-90.

Dahl, D., and Shrieves, R.E., 1990, "The impact of regulation on bank equity infusion", *Journal of Banking and Finance*, 14, 1209-1228.

Daniel, D.L., Longbrake, W.A., and Murphy, N.B., "The effect of technology on bank economies of scale for demand deposits", *Journal of Finance*, 28, 131-146.

Demsetz, H., 1973, "Industry structure, market rivalry, and public policy", *Journal of Law and Economics*, 16(1): 1-9.

Dermine, J., and Hillion, P., 1992, "Deposit rate ceilings and the market values of banks: The case of France, 1971-1981", *Journal of Money, Credit and Banking*, 24(2): 184-194.

Dewatripont, M., and Tirole, J., 1993, "Efficient governance structure: Implications for banking regulation", in C. Mayer and X. Vives (eds), *Capital Markets and Financial Intermediation*, Cambridge: Cambridge University Press, 12-33.

DeYoung, R., and Hasan, I., 1998, "The performance of de novo commercial banks: A profit efficiency approach", *Journal of Banking and Finance*, 22, 565-587.

Diamond, D.W., 1984, "Financial intermediation and delegated monitoring", *Review of Economic Studies*, 51(3): 393-414.

Diamond, D.W., 1993, "Bank loans maturity and priority when borrowers can refinance", in C. Mayer and X. Vives (eds), *Capital Markets and Financial Intermediation*, Cambridge: Cambridge University Press, 46-65.

Diamond, D.W., 1996," Financial intermediation as delegated monitoring: A simple example", Federal Reserve Bank of Richmond *Economic Quarterly*, 82(3), 51-67.

Diamond, D.W., and Dybving, P.H., 1983, "Bank runs, deposit insurance, and liquidity", *Journal of Political Economy*, 91(3): 401-419.

Diamond, D.W., and Dybving, P.H., 1986, "Banking theory, deposit insurance, and bank regulation", *Journal of Business*, 59(1): 55-69.

Diamond, D.W., and Rajan, R.G., 1998, "Liquidity risk, liquidity creation and financial fragility: A theory of banking", Working Paper, University of Chicago.

Dietsch, M., 1988, "Economies d'echelle et economies d' envergure dans les banques de depots francaises'", Mimeo, Instutut d' Etudes Politiques de Strasbourg.

Dietsch, M., 1993, "Economies of scale and scope in French commercial banking industry", *Journal of Productivity Analysis*, 4(1-2): 35-50.

Dietsch, M., and Vivas, A.L., 1996, "How the environment determines the efficiency of banks: A comparison between French and Spanish banking industry", Working Paper.

Diewert, W.E., 1974, "Application of duality theory", in M.D. Intriligator and D.A. Kendrick (eds), *Frontiers of Quantitative Economics*, Vol. II, Amsterdam: North-Holland.

Diewert, W.E., 1982, "Duality approaches to microeconomic theory", Chapter 12 in K.J. Arrow and M.D. Intriligator (eds), *Handbook of Mathematical Economics*, Vol. II, Amsterdam: North-Holland.

Dowd, K., 1994, "Competitive banking, bankers' club, and banking regulation", *Journal of Money, Credit and Banking*, 26(2): 289-308.

Drake, L., 1992, "Economies of scale and scope in UK building societies: an application of the translog multiproduct cost function", *Applied Financial Economics*, 2, 211-219.

Drake, L., 1995, "Testing for expense preference behaviour in UK building societies", *The Service Industries Journal*, 15, 50-65.

Drake, L., and Weyman-Jones, T.G, 1992, "Technical and scale efficiency in the UK building societies", *Applied Financial Economics*, 2, 1-9.

Ediz, T., Micheal, I., and Perraudin, W., 1998, "The impact of capital requirements on UK bank behaviour", Federal Reserve Bank of New York *Economic Review*, 4(3): 15-23.

Edwards, F.R., 1964, "Concentration in banking and its effects on business loan rates", *Review of Economics and Statistics*, 46(3): 294-300.

Edwards, F.R., 1973, "Advertising and competition in banking", *Antitrust Bulletin*, 18(1): 23-32.

Elyasiani, E., and Mehdian, S., 1990, "Efficiency in the commercial banking industry, a production frontier approach ", *Applied Economics*, 22, 539-551.

Elyasiani, E., and Mehdian, S., 1992, "Productive efficiency performance of minority and nonminority-owned banks: A non-parametric approach", 1992, *Journal of Banking and Finance*, 16, 933-948.

Elyasiani, E., and Mehdian, S., 1995, "The comparative efficiency performance of small and large US commercial banks in the pre- and post-deregulation eras", *Applied Economics*, 27, 1069-1079.

Elyasiani, E., Mehdian, S., and Rezvanian, R., 1994, "An empirical test of association between production and financial performance: the case of commercial banking industry", *Applied Financial Economics, 4*, 55-59.

English, M, Grosskopf, S., Hayes, K, and Yaisawarng, S., 1993, "Output allocative and technical efficiency of banks", *Journal of Banking and Finance*, 17, 349-366.

European Economy, 1988, 'The economics of 1992', No.32, March.

Evanoff, D.D., and Fortier, D.L., 1988, "Reevaluation of the structure-conduct-performance paradigm in banking", *Journal of Financial Services Research*, 1(3): 277-94.

Eyssell, T., and Arshadi, N., 1990, "The wealth effects of the risk-based capital requirement in banking", *Journal of Banking and Finance*, 14, 179-197.

Fama, E., 1985, "What's different about banks? ", *Journal of Monetary Economics*, 15, 29-36.

Fama, E.F., 1980, "Banking in the theory of finance", *Journal of Monetary Economics*, 6(1): 39-57.

Fare, R, Grosskopf, S., and Lovell, C.A.K., 1983, "The structure of technical efficiency", *Scandinavian Journal of Economics*, 85(2): 181-190.

Fare, R, Grosskopf, S., and Lovell, C.A.K., 1985, *The Measurement of Efficiency of Production*, Boston, Mass: Kluwer - Nijhoff Publishing.

Fare, R, Grosskopf, S., and Lovell, C.A.K., 1994, *Production Frontiers*, Cambridge: Cambridge University Press.

Fare, R., and Primont, D., 1993, "Measuring the efficiency of multiunit banking: An activity analysis approach", *Journal of Banking and Finance*, 17, 539-544.

Farell, M.J., 1957, "The measurement of productive efficiency", *Journal of Royal Statistical Society*, 120(3): 253-281.

Farrier, G.D, and Lovell, CA.K., 1990, "Measuring cost efficiency in banking: Econometric and linear programming evidence", *Journal of Econometrics*, 46, 229-245.

Favero, C.A., and Luca, P. 1993, "Technical and scale efficiency in the Italian banking sector. A non-parametric approach", Working Paper: Universita Cattolica del Sarco Cuore, no. 11.

Fields, J.A., Murphy, N.B., and Tirtiroglou, D., 1993, "An international comparison of scale economies: Evidence from Turkey", *Journal of Financial Services, Research*, 7, 111-125.

Financial Services Authority, Annual Report, 1997/1998.

Financial Services Authority, *Annual Report*, 1998/1999.

Financial Services Authority, *Summary Annual Report*, 1998/1999.

Fisher, F.M., and McGowan, J.J., 1986, "On the misuse of accounting rates of return to infer monopoly profits", *American Economic Review*, 73(1): 82-97.

Fleschig, T.G., 1965, "The effects on concentration on bank loan rates", *Journal of Finance*, 17, 437-450.

Fraser, D., and Rose, P.S., 1971, " More on banking structure and performance: The evidence from Texas", *Journal of Financial and Quantitative Analysis*, 9, 287-296.

Fraser, D.R., and Alvis, J.B., 1975, "The structure-performance relationship in banking, a dichotomous analysis", *Review of Business and Economic Research*, 2, 35-57.

Fraser, D.R., and Rose, P. S., 1976, "Static and dynamic measures of market structure and the performance of commercial banks", *Journal of Economics and Business*, 28(2): 79-87.

Freeman, S., 1988, "Banking as the provision of liquidity", *Journal of Business*, 61(1): 45-64.

Freixas, X., and Rochet, J.C, 1998, *Microeconomics of Banking*, US: The MIT Press, Second edition.

Furlong, F., and Keely, M., "Capital regulation and bank risk-taking", *Journal of Banking and Finance*, 15(4/5): 1523-1542.

Fuss, R.A., and McFadden, D., 1980, *Production Economics: A Dual Approach to Theory and Application*, Amsterdam: North-Holland.

Gardener, E.P.M., and Molyneux, P., 1994, *Changes in Western European Banking*, UK: Cambridge University Press.

Gennotte, G., and Pyle, D., 1991, "Capital controls and bank risk", *Journal of Banking and Finance*, 15, 805-824.

Geroski, P.A., 1991, *Market Dynamic and Entry*, Oxford: Blackwell.

Giammarino, R.M., Lewis, D.E., and Sappington, D.E.M., 1993, "An incentive approach to banking regulation", *Journal of Finance*, 48(4): 1523-1542.

Gilbert, R.A., 1984, "Bank market structure and competition", *Journal of Money, Credit and Banking*, 16(4): 617-645.

Gilligan, T.W. and Smirlock, M.L., 1984, "An empirical study of joint production and scale economies in commercial banking", *Journal of Banking and Finance*, 8, 67-77.

Gjerde, O. and Semmen, K., 1995, "Risk-Based Capital Requirements and Bank Portfolio Risk", *Journal of Banking and Finance*, 19(7): 1159-73.

Glass, J.C., and McKillop, D.C., 1992, "An empirical analysis of scale and scope economies and technical changes in an Irish multiproduct banking firm", *Journal of Banking and Finance*, 16(2): 423-37.

Glassman, C.A., and Rhoades, S.A., 1980, "Owner vs. manager control effects on bank performance", *Review of Economics and Statistics*, 62(2): 263-270.

Goldberg, L.G., and Rai, A., 1994, "The impact of risk-based capital standards on World banks", *Journal of International Financial Markets, Institutions, and Money*, 3(3-4): 85-100.

Goldberg, L.G., and Rai, A., 1996, "The structure-performance relationship for European banks", *Journal of Banking and Finance*, 20, 745-771.

Gorton, G, and Pennacchi, G., 1990, "Financial intermediaries and liquidity creation", *Journal of Finance*, 45(1): 49-72.

Grabowski, R., Rangan, N., and Rezvanian, R., 1993, "Organisational forms in banking: An empirical investigation of cost efficiency", *Journal of Banking and Finance*, 17, 531-538.

Grabowski, R., Rangan, N., and Rezvanian, R., 1994, "The effect of deregulation on the efficiency of US banks", *Journal of Economics and Business*, 46(1): 39-54.

Graddy, D.B., and Kyle, R., 1979, "The simultaneity of bank decision-making, market structure, and bank performance", *Journal of Finance*, 34(1): 1-18.

Gramley, L.E., 1962, "A study of scale economies in banking", Working paper, Federal Reserve Bank of Kansas City.

Green, W.H., 1990, "A gamma-distributed stochastic frontier model", *Journal of Econometrics*, 46, 141-163.

Green, W.H., 1993, "The econometric approach to efficiency analysis", in M.O. Fried, C.A.K. Lovell and S.S. Schmidt (eds), *The Measurement of Productive Efficiency: Techniques and Applications*, Oxford: Oxford University Press.

Green, W.H., 1995, *LIMDEP*, Australia: Econometric Software, Inc.

Greene, W.H., 1996, *Econometric Analysis*, USA: Macmillan Publishing Company, Second Edition.

Gropper, D.M., 1991, "An empirical investigation of changes in scale economies for commercial banking firm, 1976-1986", *Journal of Money, Credit and Banking*, 23(4): 718-727.

Gual, J. and Neven, D., 1993, "Deregulation of the European banking industry", *European Economy*, 3, 151-182.

Gurley, J., and Shaw, E., 1960, *Money in the Theory of Finance,* Washington: Brookings Institution.

Hadjiemmanuil, C. 1996, *Banking Regulations and the Bank of England,* London: LLP Limited.

Hancock, D., 1985, "Bank profitability, interest rates and monetary policy", *Journal of Money, Credit and Banking,* 17(2): 189-202.

Hancock, D., 1991, *A Theory of Production for the Financial Firm*, Norwell, Mass: Kluwer Academic Publishers.

Hancock, D., and Wilcox, J.A., 1992, "The effects on bank assets of business conditions and capital shortfalls", *Proceedings of a Conference on Bank Structure and Competition*, Federal Reserve Bank of Chicago, 502-520.

Hannan, T., 1979, "Limit pricing and the banking industry", *Journal of Money, Credit, and Banking,* 11(4): 438-446.

Hannan, T., 1983, "Bank profitability and the threat of entry", *Journal of Bank Research*, 14(2): 157-63.

Hardwick, P., 1989, "Economies of scale in building societies", *Applied Economics*, 21, 1291-1304.
Hardwick, P., 1989, "Economies of scale in building societies", *Applied Economics*, 21, 1291-1304.
Haubrich, J.G., and Wachtel, 1993, "Capital requirements and shifts in commercial bank portfolio", Federal Reserve Bank of Cleveland *Economic Review*, 29(3): 2-15.

Hay, D.A., and Morris, D.J., 1979, *Industrial Economics: Theory and Evidence*, Oxford: Oxford University Press.

Heffernan, S. 1996, *Modern Banking in Theory and Practice*, West Sussex: John Wiley & Sons Ltd.

Heggestad, A.A., 1977," Market structure, risk and profitability in commercial banking", *Journal of Finance*, 32(4): 1207-1216.

Heggestad, A.A., and Mingo, J. J., 1976, "Prices, non-prices, and concentration in commercial banking", *Journal of Money, Credit, and Banking,* 8(1): 107-117.

Hellwig, M., 1991; "Banking, financial intermediation and corporate finance", in A. Giovanninin and C. Mayer (eds), *European Financial Integration*, Cambridge: Cambridge University Press, 35-57.

Hicks, J.R., 1935, "Annual survey of economic theory: The theory of monopoly", *Econometrica*, 3, 1-20.

Honohan, P., and Kinsella, R.P., 1982, "Comparing Bank Concentration Across Countries", *Journal of Banking and Finance*, 6, 255-262.

Horvitz, P.M., 1963, "Economies of scale in banking", in P. Horvitz et al., *Private Financial Institutions*, Englewood Cliffs, New Jersey: Prentice-Hall, 1-25.

Humphrey, D.B., 1993, "Cost and technical change: Effects from bank deregulation", *Journal of Productivity Analysis*, 4(1-2): 9-34.

Humphrey, D.B., and Pulley, L.B., 1997, "Banks' responses to deregulation: Profits, technology, and efficiency", *Journal of Money, Credit and Banking*, 29(1): 73-93.

Intriligator, M., Bodkin, R, and Hsiao, C., 1996, *Econometric Models, Techniques, and Applications*, USA: Prentice-Hall, Inc., Second Edition.

Jackson, W.E., 1992, "The price-concentration relationship in banking: A comment", *Review of Economics and Statistics*, 74, 373-376.

Jacobs, D.P., 1971, Business Loan Costs and Bank Market Structure, *National Bureau of Economic Research Occasional Paper*, No. 15.

Jagtiani, J., and Khanthavit, A., 1996, "Scale and scope economies at large banks: Including off-balance sheet products and regulatory effect (1984-1991)", *Journal of Banking and Finance*, 20, 1171-1287.

Jagtiani, J., Nathan, A., and Sick, G., 1995 "Scale economies and cost complementarities in commercial banks: On- and off-balance-sheet activities", *Journal of Banking and Finance*, 19, 1175-1189.

James, A.S., and Richardson, T., "The underwriting experience of commercial bank affiliates prior to the Glass-Steagal Act: A re-examination of evidence for passage of the Act", *Journal of Banking and Finance*, 18(2): 351-395.

James, C., 1983, "An analysis of intra-industry differences in the effect of regulation: The case of deposit rate ceilings", *Journal of Monetary Economics*, 12, 417-432.

James, C., 1987, "Some evidence on the uniqueness of bank loans", *Journal of Financial Economics*, 19, 217-235.

Jayarante, J. and Strahan, P.E., 1996, "The finance-growth nexus: Evidence from bank branch deregulatio", *Quarterly Journal of Economics*, 111(3): 639-670.

Jayarante, J. and Strahan, P.E., 1999, "The effects of Regulation on Bank Performance in the United States", in P. Harker and S.A. Zenios (eds), *Performance of Financial Institutions: Efficiency, Innovation, Regulation*, Cambridge University Press.

Jondrow, J., Lovell, C.A., Meterov, I.S., and Schmidt, P., 1982, "On estimation of technical inefficiency in the stochastic frontier model", *Journal of Econometrics*, 19, 233-238.

Kalish, L., and Gilbert, R.A., 1973, "The influence of bank regulation on the operating efficiency of commercial banks", *Journal of Finance*, 28(5): 1287-1301.

Kaparakis, E.I., Miller, S.M., and Noulas, A.G., 1994, "Short-run cost inefficiency of commercial banks: A flexible stochastic frontier approach", *Journal of Money, Credit and Banking*, 26(1): 875-893.

Keely, M.C., and Furlong, 1990, "A re-examination of mean variance analysis of bank capital regulation", *Journal of Banking and Finance*, 14, 69-84.

Kenneth, S.B., and Madura, J., 1996, "Wealth effects of reserve requirement reductions in the 1990s on depository institutions", *Review of Financial Economics*, 5(2): 191-204.

Kim, D., and Santomero, A., 1988, "Risk in banking and capital regulation", *Journal of Finance*, 43, 1219-1233.

Kinsella, R.P., 1980, "The measurement of bank output", *Journal of the Institute of Bankers in Ireland*, 82, 173-183.

Klausner, M., and White, L.J., 1993, "Bank Regulatory Reform and Bank Structure: Introduction", in M. Klausner, and L.J. White (eds), *Structural Change in Banking*, USA: New York University Salomon Centre, 1-18.

Klein, M.A., 1971, "A theory of a banking firm", *Journal of Money, Credit and Banking"*, 3, 205-218.

Koehn, M., and Santomero, A., 1980, "Regulation of bank capital and portfolio risk", *Journal of Finance*, 35, 1235-1250.

Koehn, M., and Stangle, B., 1980," The effect if deposit-rate ceilings on bank risk: A comment", *Journal of Banking and Finance*, 4, 381-386.

Kolari, J. and Zardakoohi, 1987, A., *Bank Costs, Structure and Performance*, Lexington, Mass: D.C: Heath.

Kolari, J., Mahajan, A., and Saunders, E., 1988, "The effect of changes in reserve requirements on bank stock prices", *Journal of Banking and Finance*, 12, 184-198.

Kroszner, R.A., and Razan, R.G., 1994, "Is the Glass-Steagall Act justified? A study if the US experience with universal banking before 1933", *American Economic Review*, 84(4): 810-832.

Kroszner, R.A., and Razan, R.G., 1997, "Organisation structure and credibility: Evidence from commercial bank securities activities before the Glass-Steagall Act", *Journal of Monetary Economics*, 39(3): 475-516.

Kumbhakar, S.C., 1996, "A parametric approach to efficiency measurement using a flexible profit function", *Southern Economic Review*, 63(2): 473-487.

Kupiec, P.H., and O'Brien, J.M., 1998, "Deposit insurance, bank incentives, and the design of regulatory policy", Working Paper, *Board of Governors of the Federal Reserve System*, 1998-10.

Kwast, M.L., and Rose, T., 1982, "Pricing, operating efficiency, and profitability among large commercial banks", *Journal of Banking and Finance*, 6, 233-254.

Lang, G., and Welzer, P., 1996, "Efficiency and technical progress in banking: Empirical results from a panel of German co-operative banks", *Journal of Banking and Finance*, 20, 1003-1023.

Lawrence, C. and Shay, R., 1986, "Technology and financial intermediation in multiproduct banking firm: an econometric study of I.S.S. banks, 1979-1982", in C. Lawrence and R. Shay (eds), *Technological Innovation, Regulation and the Monetary Economy*, Cambridge, Mass: Ballinger.

Leite, A.N., and Barros, P.P., 1994, "Conduct and Competition in Portuguese commercial banking", Working Paper, Universidade Nova De Lisboa, 216.

Leland H., and Pyle, D., 1977, "Information asymmetries, financial structure and financial intermediaries", *Journal of Finance*, 32(2): 371-87.

Lewis, M.K., and Davis, K.T., 1987, *Domestic and International Banking*, UK: Philip Allan.

Libenstein, H., 1966, "Allocative efficiency vs. 'X-efficiency'", *American Economic Review*, 56, 392-415.

MacDonald, R., 1996, "Capital Requirements", *Handbooks in Central Banking*, Centre for Banking Studies, Bank of England.

MacDonald, R., 1996, "Deposit Insurance", *Handbooks in Central Banking*, Centre for Banking Studies, Bank of England.

Mansur, I., and Elyasiani, E., 1994, "An examination of the impact of the 1989 FIRREA on the market value of commercial banks and savings and loans", *Applied Financial Economics*, 4, 11-22.

Marlow, M.L., 1982, "Bank structure and mortgage rates: Implications for interstate banking", *Journal of Economics and Business*, 34(2): 135-42.

Martin, F., and Sassenou, M., 1994, "Cost structure in French banking: A re-examination based on a regular CES-Quadratic form" *Journal of International Financial Markets, Institutions and Money,* 3(3-4), 137-69.

Matutes, C., and Vives, X., 1996, "Competition for deposits, fragility, and insurance", *Journal of Financial Intermediation*, 5(2): 184-216.

Matutes, C., and Vives, X., 1996, "Imperfect competition, risk taking, and regulation in banking", *Centre for Economic Policy Research*, Discussion Paper, 1177.

Mayer, C., 1993, "European banking: Prudential and regulatory issues", in J. Dermine (eds), *European Banking in the 1990s*, Oxford: Blackwell Publishers, Second Edition, 43-64.

Mayer, C., and Neven, D., 1991, "European financial regulation: A framework for policy analysis", in A. Giovannini and C. Mayer (eds), *European Financial Integration*, Cambridge: Cambridge University Press, 112-129.

Mayer, P.A., 1967, "Price discrimination, regional loan rates and the structure of the banking industry", *Journal of Finance*, 22, 37-48.

Mayer, C., and Vives, X., 1993, "Capital markets and financial intermediation: Introduction", in C. Mayer and X. Vives (eds), *Capital Markets and Financial Intermediation*, Cambridge: Cambridge University Press, 1-11.

McDonald, R., 1996, "Deposit Insurance", in *Handbooks in Central Banking*, Centre for Banking Studies, No.9.

McKillop, D.G., and Glass, C.J, 1994, "A cost model of building societies as producers of mortgages and other financial products, *Journal of Business, Finance and Accounting*, 21(7): 1031-1046.

McKillop, D.G., Glass, J.C., and Morikawa, Y., 1996, "The composite cost function and efficiency in giant Japanese banks", *Journal of Banking and Finance*, 20, 1651-1671.

Merrick, J.J., and Saunders, A., 1985, "Bank Regulation and Monetary Policy", *Journal of Money, Credit and Banking*, 17(2): 693-726.

Mester, L.J., 1992, "Traditional and nontraditional banking: An information - theoretic approach", *Journal of Banking and Finance*, 16(2): 454-466.

Mester, L.J., 1993, "Efficiency in the saving and loan industry", *Journal of Banking and Finance*, 17, 267-286.

Mester, L.J., 1996, "A study of bank efficiency taking into account risk-preferences" *Journal of Banking and Finance*, 20(6): 1025-1045.

Miller, S.M., and Noulas, A.G., 1996, "The technical efficiency of large bank production", *Journal of Banking and Finance*, 20, 495-509.

Mingo, J., 1976, "Managerial motives, market structures and the performance of holding company banks", *Economic Inquiry*, 14(3): 411-424.

Mingo, J., 1978, "The effect of deposit rate ceilings on bank risk", *Journal of Banking and Finance*, 2, 367-378.

Mingo, J., 1980, "The microeconomics of deposit rate ceilings: Inference for NOW accounts and interest on checking accounts", *Journal of Banking and Finance*, 4, 387-395.

Mitchell, K., and Onvural, N.M., 1996, "Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form", *Journal of Money, Credit and Banking*, 28(2): 178-199.

Molyneux, P., Altunbas, Y, and Gardener, E, 1996, *Efficiency in European Banking*, UK: John Wiley & Sons Ltd.

Molyneux, P., and Forbes, W., 1995, "Market structure and performance in European Banking", *Applied Economics*, 27, 155-159.

Molyneux, P., and Thorton, J., 1992, "Determinants of European bank profitability: A note", *Journal of Banking and Finance*, 16, 1173-1178.

Mullineaux, D.J, 1975, "Economies of scale of financial institutions", *Journal of Monetary Economics*, 1, 233-240.

Murphy, N.B., 1972, "A re-estimation of the Benston-Bell Murphy cost function for larger sample with greater size and geographic dispersion", *Journal of Quantitative Analysis*, 7, 2097-2105.

Murphy, N.B., 1972, "Cost of banking activities: interaction between risk and operation cost: a comment", *Journal of Money, Credit, and Banking*, 4, 614-615.

Nadiri, M.I., 1982, "Producers theory", Chapter 10 in K.J. Arrow and M.D. Intrtiligator (eds), *Handbook of Mathematical Economics*, Vol. II, Amsterdam: North-Holland, 431-490.

Nakamura, L.I., 1993, "Commercial bank information: Implications for the structure of banking", in M. Klausner., and L.J. White (eds), *Structural Change in Banking*, USA: New York University Salomon Centre, 131-160.

Nathan, A., and Neave, E.H., 1992, "Operating Efficiency of Canadian Banks", *Journal of Financial Services Research*, 6, 265-276.

Nerlove, M., 1963, "Returns to Scale in the Electricity Supply" in C. Crist (ed.) *Measurements in Economics*, California: Standford University Press.

Nerlove, M., 1963, "Returns to Scale in the Electricity Supply", in *Measurements in Economics*, California: Standford University Press, p.p. 167-98.

Noulas, A.G., Miller, S.M., and Ray, S.C., 1993, "Regularity conditions and scope estimates: The case of large-sized US banks", *Journal of Financial Services Research*, 235-248.

Osborne, D.K., and Wendel, J., 1981, "A note on concentration and checking account prices", *Journal of Finance*, 36(1): 181-186.

Osborne, D.K., and Zaher, T.S., 1992, "Reserve requirements, bank share prices, and the uniqueness of bank loans", *Journal of Banking and Finance*, 16, 799-812.

Ostrosky, A.L., 1997, "Bank failure in the US: A note on Regulation Q", *International Advances in Economic Research*, 3(2): 176-180.

Pappas, J.L, Brigham, E.F., and Shipley, B., 1991, *Managerial Economics*, New York: The Dryden Press.

Partington, I, 1989, *Applied Economics in Banking and Finance*, UK: Oxford University Press, Fourth Edition.

Peltzman, S., 1976, "Towards a more general theory of regulation", *Journal of Law and Economics*, 19(2): 211-240.

Peltzman, S., 1977, "The gains and losses from industrial concentration", *Journal of Law and Economics*, 20, 229-263.

Phillips, A., 1967, "Evidence on concentration in banking markets and interest rates", *Federal Reserve Bulletin*, 53(6): 916-926.

Pi, L., and Timme, S.G., 1993, "Corporate control and bank efficiency", *Journal of Banking and Finance*, 17, 515-530.

Pindyck, R.S., and Rubinfeild, D.L., 1989, *Microeconomics*, Singapore: Macmillan Publishing Company.

Pindyck, R.S., and Rubinfeild, D.L., 1991, *Econometric Models & Economic Forecasts*, Singapore: McGraw-Hill International Editions.

Preece, C.P., and Mullineux, A.J., 1994, "Monitoring by financial intermediaries: Banks vs. nonbanks", *Journal of Financial Service Research*, 8, 193-202.

Pulley, L. and Humphrey, D.B., 1993, "The Role of Fixed Costs and Cost Complementarities in Determining Scope Economies and the Cost of Narrow Banking Proposals", *Journal of Business*, 66(3): 437-62.

Pulley, L.B., and Braunstein, Y.M., 1992, "A composite cost function for multiproduct firms with an application to economies of scope in banking", *Review of Economics and Statistics*, 74, 221-230.

Puri, M., 1994, "The long term default performance of bank underwritting security issues", *Journal of Banking and Finance*, 18(2): 397-418.

Puri, M., 1996, "Commercial banks in investment banking: Conflict of interest or certification role?", *Journal of Financial Economics*, 40(3): 373-401.

Pyle, D., 1974, "Loss of income on savings deposits from interest rate regulations", *The Bell Journal*, 5(2): 114-122.

Rhoades, S.A., 1980, "Monopoly and expense preference behaviour: An empirical investigation of a behavioralist hypothesis", *Southern Economic Journal*, 47(2): 419-432.

Rhoades, S.A., 1981, "Does market structure matter in commercial banking?", *Antitrust Bulletin,* 26(1): 155-181.

Rhoades, S.A., 1982a, "Welfare loss, redistribution effect, and restriction of output due to monopoly in banking", *Journal of Monetary Economics*, 9(3): 375-387.

Rhoades, S.A., 1982b, "The relative size of banks and industrial firms in the United States and other countries- a note", *Journal of Banking and Finance*, 6, 579-585.

Rhoades, S.A., and Rutz, R.D., 1982, "Market power and firm risk: A test of the 'Quiet Life' hypothesis", *Journal of Monetary Economics*, 9(1): 73-85.

Rhoades, S.A., and Savage, D.T., 1981, "The relative performance of bank holding companies and branch banking systems", *Journal of Economics and Business,* 33(2): 132-141.

Rochet, J.C., 1992, "Capital requirements and the behaviour of commercial banks", *European Economic Review*, 36(5): 1137-1170.

Rogers, K.E, 1998a, "Product mix, bank powers, and complementarities at US commercial banks", *Journal of Economics and Business*, 50, 205-218.

Rogers, K.E, 1998b, "Nontraditional activities and the efficiency of US commercial banks", *Journal of Banking and Finance*, 22, 467-482.

Rose, P.S., 1987, "The impact of financial service deregulation: The hypothesis and the evidence from 240 US metropolitan banking markets", *Quarterly Journal of Business and Economics*, 26, 55-88.

Rose, P.S., and Scott, W.L., 1979, "The performance of banks acquired by holding companies", *Review of Business and Economic Research*, 14(3): 18-37.

Salomon Brothers, 1990, *Multinational Money Centre Banking: The Evaluation of a Single European Market*, London: Salomon Brothers.

Santomero, A.M., 1984, "Modelling the banking firm: A survey", *Journal of Money, Credit and Banking*, 16(4), 576-601.

Santomero, A.M., 1997, "Deposit insurance: Do we need it and why?", Working Paper, The Wharton Financial Institutions Centre, University of Pennsylvania, 97-35.

Savage, D.T., and Rhoades, S.A., 1979, "The effect of branch banking on pricing, profits and efficiency of unit banks", in *Bank Structure and Competition, Proceedings of a Conference at the Federal Reserve Bank of Chicago*, 187-196.

Schaefer, S.M., 1987, "The design of bank regulation & supervision: Some lessons from the theory of finance", Working Paper, London Business School.

Scherer, F.M., 1980, *Industrial Market Structure and Economic Performance*, Chicago: Rand McNally, Second Edition.

Schmidt, P., 1985, "Frontier production functions", *Econometric Reviews*, 4(2): 289-328.

Schweiger, I., and McGee, J.S., 1961, "Chicago banking", *Journal of Business*, 34(3): 203-366.

Seiford, L.M., and Thorall, R.M., 1990, "Recent developments in DEA: The mathematical programming approach to frontier analysis", *Journal of Econometrics*, 46, 7-38.

Shephard, R.W., 1953, *Cost and Production Functions*, Princeton, NJ: Princeton University Press.

Shephard, R.W., 1970, *The Theory of Cost and Production Function*, Princeton, NJ: Princeton University Press.

Shepherd, W.G. 1982, "Economies of scale and monopoly profits", in J.V. Craven (ed) *Industrial Organisation, Antitrust, and Public Policy*, Boston: Kluwer Nijhoff.

Shepherd, W.G., 1986, "Tobin's q and the structure-performance relationship: Comment", *American Economic Review*, 76(5): 1205-1210.

Shepherd, W.G., 1990, *The Economics of Industrial Organisation*, USA: Prentice-Hall, Inc., Third Edition.

Sherman, H.D., and Gold, F., 1985, "Bank branch operating efficiency", *Journal of Banking and Finance*, 9, 297-315.

Shiers, A.F., 1994, "Deposit insurance and banking system risk: Some empirical evidence", *Quarterly Review of Economics and Finance*, 34(4): 347-361.

Shrieves, R.E., and Dahl, D., 1992, "The relationship between risk and capital in commercial banks", *Journal of Banking and Finance*, 11, 581-600.

Slovin, M., Sushka, E., and Bendeck, Y., 1990, "The market valuation effects of reserve regulation", *Journal of Monetary Economics*, 25, 451-473.

Smirlock, M., 1984, "An analysis of bank risk and deposit rate ceilings: Evidence from the capital market", *Journal of Monetary Economics*, 13, 195-210.

Smirlock, M., 1985, "Evidence on the (non) relationship between concentration and profitability in Banking", *Journal of Money, Credit and Banking*, 17(1): 69-83.

Smirlock, M., Gilligan, T., and Marshall, W., 1984, "Tobin's q and the structure-performance relationship", *American Economic Review*, 74, 1050-1060.

Smirlock, M., Gilligan, T., and Marshall, W., 1986, "Tobin's q and the structure- performance relationship: A reply", *American Economic Review*, 76, 1211-1213.

Soteriou, A., and Zenios, S.A., 1997, "Efficiency, Profitability and quality of banking services", Working Paper, The Wharton Financial Institutions Centre, University of Pennsylvania, 97-28.

Startz, R., 1979, "Implicit interest on demand deposits", *Journal of Monetary Economics*, 5, 515-534.

Stevenson, R.E., 1980, "Likelihood functions for generalised stochastic frontier estimation", *Journal of Econometrics*, 13, 57-66.

Syron, R.F., and Randall, R.E., 1992, "The procyclical application of bank capital requirements", Federal Reserve Bank of Boston *1991Annual Report*.

Thakor, A.V., 1993, "Capital requirements, monetary policy, and aggregate bank lending" Indiana University Working Paper, August.

Thakor, A.V., 1996, "The design of the financial systems: An overview", *Journal of Banking and Finance*, 20, 917-948.

Theil, H., 1971, *Principles of Econometrics*, US: John Wiley & Sons, Inc.

Thomas, R.L., 1993, *Introductory Econometrics: Theory and Applications*, UK: Longman Group Ltd, Second Edition.

Tirole, J., 1988, *The Theory of Industrial Organisation*, Cambridge, Mass: The MIT Press.

Tirole, J., 1994, "On banking and intermediation", *European Economic Review*, 38(3-4): 467-87.

Vernon, J.R., 1971, "Separation of ownership and control and profit rates, the evidence from banking: Comment", *Journal of Financial and Quantitative Analysis*, 6, 615-625.

Vives, X., 1991, "Banking competition and European integration", in A. Giovannini and C. Mayer (eds), *European Financial Integration*, Cambridge: Cambridge University Press, 9-25.

Von Thadden, E.L., 1990, "Bank finance and long-term investment", Working Paper, *University of Basel.*

Walter, I., and Smith, R.C., 1993, "European investment banking: Structure, transactions flow and regulation", in J. Dermine (ed.), *European Banking in the 1990s*, Oxford: Blackwell Publishers, Second Edition, 109-156.

Ware, R.F., 1972, "Banking structure and performance: Some evidence from Ohio", Working Paper, Federal Reserve Bank of Cleveland Economic Review, March, 3-14.

Wheelock, D.C., and Wilson, P.W., "Deposit insurance, regulation, and efficiency", *Review of Economics and Statistics*, 77(4): 689-700.

White, L.J., 1999, "Technological changes, financial innovation, and financial regulation in the US: The challenges for public policy", in P. Harker and S.A. Zenios (eds), *Performance of Financial Institutions: Efficiency, Innovation, Regulation*, Cambridge University Press.

Whitehead, D.D., 1977, "Holding company power and market performance: A new index of market concentration", Working Paper, Federal Reserve Bank of Atlanta, December.

Whitehead, D.D., 1978, "An empirical test of the linked oligopoly theory: An analysis of Florida Holding Companies", Working Paper, Federal Reserve Bank of Atlanta, June.

Winton, A., 1995, "Delegated monitoring and bank structure in finite economy", *Journal of Financial Intermediation*, 4, 158-187.

Yates, A.J., 1974, "Further evidence on the structure-performance relationship in banking", *Journal of Economics and Business*, 26, 95-100.

Zaim, O., 1995, "The Effect of Financial Liberalization on the Efficiency of Turkish Commercial Banks", *Applied Financial Economics*, 5(4): 257-64.

Zellner, A., 1962, "An efficient method of estimating seemingly unrelated regression and test for aggregation bias", *Journal of American Statistical Association*, 57, 348-368.